

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 6, 2001, 20:59:36 : Search time 51.66 Seconds  
(without alignments)  
726.850 Million cell updates/sec

Title: US-09-292-862-2

Perfect score: 553  
Sequence: 1 MQARYSVSPNSLGVVPLG.....PSSQSLYRTSGAFVYDCSKF 553

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 195891 seqs, 6790065 residues

Word size: 0

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

1: PIR66:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	109	19.7	117	2	149674	forkhead transcrip
2	106	19.2	106	2	S51626	FRFAC-3 protein -
3	61	11.0	461	2	S34472	MFH-1 protein - mo
4	59	10.7	101	2	I60918	brain factor-3 - r
5	39	7.1	108	2	B48924	forkhead transcrip
6	35	6.3	111	2	B56556	fork head domain
7	25	4.5	508	2	S59870	forkhead transcrip
8	23	4.2	111	2	C47746	forkhead transcrip
9	22	4.0	111	2	F47746	hypothetical prote
10	21	3.8	270	2	T16880	SLP2 protein - fru
11	19	3.4	445	2	S23055	SLP2 protein - fru
12	19	3.4	445	2	S23056	SLP2 protein - fru
13	16	2.9	106	2	S51630	SLP2 protein - fru
14	16	2.9	109	2	B54743	SLP2 protein - fru
15	16	2.9	134	2	B46178	SLP2 protein - fru
16	16	2.9	387	2	A47446	SLP2 protein - fru
17	16	2.9	451	2	A55909	SLP2 protein - fru
18	16	2.9	469	2	A54743	SLP2 protein - fru
19	16	2.9	476	2	JH0672	SLP2 protein - fru
20	16	2.9	480	2	JH0672	SLP2 protein - fru
21	15	2.7	101	2	I60922	SLP2 protein - fru
22	15	2.7	101	2	I60917	SLP2 protein - fru
23	15	2.7	106	2	S51627	SLP2 protein - fru
24	15	2.7	128	2	C46178	SLP2 protein - fru
25	15	2.7	302	2	T30169	SLP2 protein - fru
26	15	2.7	323	2	T15311	SLP2 protein - fru
27	15	2.7	440	2	S71795	SLP2 protein - fru
28	15	2.7	465	2	G02738	SLP2 protein - fru
29	14	2.5	101	2	I60920	SLP2 protein - fru

30	14	2.5	111	2	B47746	forkhead transcrip
31	13	2.4	76	2	S51629	FRFAC-6 protein -
32	13	2.4	101	2	A47450	HNF-3/fork-head ho
33	13	2.4	101	2	I60919	HNF-3/fork-head ho
34	13	2.4	111	2	G56556	fork head homolog
35	13	2.4	322	2	S23053	sloppy paired prot
36	13	2.4	367	2	S49008	fork head protein
37	13	2.4	370	2	S49008	transcription fact
38	13	2.4	443	1	S29334	transcription fact
39	13	2.4	445	1	S31224	transcription fact
40	13	2.4	517	2	T43358	hmf-3/forkhead tra
41	13	2.4	543	2	F70726	hypothetical glyci
42	13	2.4	543	2	A41285	interleukin enhanc
43	13	2.4	586	2	Jc6500	hmf-3/forkhead tra
44	13	2.4	617	2	A56051	myocyte nuclear fa
45	13	2.4	642	2	T39607	fork head protein

#### ALIGNMENTS

##### RESULT 1

149674 forkhead transcription factor fkh-1 - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 02-Jul-1996 #sequence-revision 02-Jul-1996 #text-change 24-Sep-1999

C:Accession: 149674; A47746; S36070

R:Sasaki, H.; Hogan, B.L., 1993

Development 118, 47-59, 1993

A:Title: Differential expression of multiple fork head related genes during gastrulat

A:Reference number: 149674; MUID:93387221

A:Accession: 149674

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Restrictions: 1-117 <RES>

A:Cross-references: GB:110406; NID:9404758; PID:AA83158.1; PID:9410056

R:Kaestner, K.H.; Lee, K.H.; Schlondorff, J.; Hiemisch, H.; Monaghan, A.P.; Schutz, G

Proc. Natl. Acad. Sci. U.S.A. 90, 7628-7631, 1993

A:Title: Six members of the mouse forkhead gene family are developmentally regulated.

A:Reference number: A47746; MUID:93361500

A:Accession: A47746

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-109 <RES>

A:Cross-references: EMBL:X71939; NID:9311736; PID:CA50741.1; PID:9311737

A:Note: submitted to the EMBL data library, May 1993

C:Note: sequence extracted from NCBI backbone (NCBI:P136529)

C:Keywords: DNA binding; transcription factor

F:8-99/Domain: fork head DNA-binding domain homology <FHD>

Query Match 19.7%; Score 109; DB 2; Length 117;

Best Local Similarity 100.0%; Pred. No. 9.4e-85;

Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 71 POPDMVPPSYTALITMAIONAPPKKTITLNGIYOFIMDRPEYRDNGKQNSIRNLT 130

DB 1 POPDMVPPSYTALITMAIONAPPKKTITLNGIYOFIMDRPEYRDNGKQNSIRNLT 60

OY 131 SLNCEVAVPRDDKPKGKSYWTIDPDSYMFNGSFLRRRRFFKKDA 179

OY 131 SLNCEVAVPRDDKPKGKSYWTIDPDSYMFNGSFLRRRRFFKKDA 109

DB 61 SLNCEVAVPRDDKPKGKSYWTIDPDSYMFNGSFLRRRRFFKKDA 109

RESULT 2

S51626

FRFAC-3 protein - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 15-Jul-1995 #sequence-revision 01-Sep-1995 #text-change 24-Sep-1999

C:Accession: S51626

R:Pierrou, S.; Hellqvist, M.; Samuelsson, L.; Enerbaeck, S.; Carlsson, P.

EMBO J. 13, 5002-5012, 1994

A:Title: Cloning and characterization of seven human forkhead proteins: binding site spc  
A:Reference number: S51624; MUID:95045392

A:Accession: S51626  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-106 <PIE>  
A:Cross-references: EMBL:U13221; NID:9563161; PIDN:AAA2038.1; PID:9563162  
C:Genetics:

A:Gene: GDB:FRHL7; FREAC3  
A:Cross-references: GDB:450220; OMIM:601090  
A:Map position: 6p25-6p25  
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
F:6-97/Domain: fork head DNA-binding domain homology <FHD>

## Query Match

Best Local Similarity 19.2%; Score 106; DB 2; Length 106;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 73 PKDMVKPPYSTALITMAIQNAPDKKITLNGIYQFIMDRPEFYRDNKGWNSIRHNSL 132  
Db 1 PKDMVKPPYSTALITMAIQNAPDKKITLNGIYQFIMDRPEFYRDNKGWNSIRHNSL 60  
OY 133 NECFVKVPRDDKKPKGSGYWTLPDPSYMFENGSLRRRRRFRKKD 178  
Db 61 NECFVKVPRDDKKPKGSGYWTLPDPSYMFENGSLRRRRRFRKKD 106

## RESULT 3

S34472  
MFH-1 protein - mouse

C:Species: Mus musculus (house mouse)  
C:Date: 02-Dec-1993 #sequence\_revision 01-Sep-1995 #text\_change 24-Sep-1999  
C:Accession: S34472  
R:Miura, N.; Wanka, A.; Tohyama, M.; Tanaka, K.  
FEBS Lett. 326, 171-176, 1993  
A:Title: MFH-1, a new member of the fork head domain family, is expressed in developing  
A:Reference number: S34472; MUID:93314779  
A:Accession: S34472  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-461 <MID>  
A:Cross-references: GB:S63607; NID:9386637; PIDN:AAB27463.1; PID:9386638  
C:Genetics:  
A:Gene: MFH-1  
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
F:38-129/Domain: fork head DNA-binding domain homology <FHD>

## Query Match

Best Local Similarity 11.0%; Score 61; DB 2; Length 461;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 NKOGWNSIRHNSLNECFVKVPRDDKKPKGSGYWTLPDPSYMFENGSLRRRRRFRKK 177  
Db 78 NKOGWNSIRHNSLNECFVKVPRDDKKPKGSGYWTLPDPSYMFENGSLRRRRRFRKK 137  
OY 178 D 178  
Db 138 D 138

## RESULT 4

160918  
Brain factor-3 - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 24-Sep-1999  
C:Accession: 160918  
R:Cleveland, D.E.; Overdier, D.G.; Tao, W.; Qian, X.; Paul, L.; Lai, E.; Costa, R.H.  
Proc. Natl. Acad. Sci. U.S.A. 90, 3948-3952, 1993  
A:Title: Identification of nine tissue-specific transcription factors of the hepatocyte  
A:Reference number: A47450; MUID:93248207  
A:Accession: 160918

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-101 <RES>

A:Cross-references: GB:LI3193; NID:9310156; PIDN:AAA41320.1; PID:9310157  
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
F:3-94/Domain: fork head DNA-binding domain homology <FHD>

## Query Match

Best Local Similarity 10.7%; Score 59; DB 2; Length 101;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 NKOGWNSIRHNSLNECFVKVPRDDKKPKGSGYWTLPDPSYMFENGSLRRRRRFRKK 176  
Db 43 NKOGWNSIRHNSLNECFVKVPRDDKKPKGSGYWTLPDPSYMFENGSLRRRRRFRKK 101

## RESULT 5

B48924

forkhead transcription activator homolog (clone FKH H8) - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 12-Sep-1997  
C:Accession: B48924  
R:Thomas, R.; Moore, J.; Johnston, T.; Socha, C.; Klemenz, M.  
Blood 81, 2854-2859, 1993  
A:Title: Drosophila forkhead homologues are expressed in a lineage-restricted manner  
A:Reference number: A48924; MUID:93271467  
A:Accession: B48924  
A:Molecule type: mRNA  
A:Status: preliminary; not compared with conceptual translation  
A:Experimental source: HEL erythroleukemia cell line  
A:Note: sequence extracted from NCBI backbone (NCBIP:133164)  
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
F:9-100/Domain: fork head DNA-binding domain homology <FHD>

## Query Match

Best Local Similarity 7.1%; Score 39; DB 2; Length 108;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 113 PFYRDNKGWNSIRHNSLNECFVKVPRDDKKPKGSGY 151  
Db 44 PFYRDNKGWNSIRHNSLNECFVKVPRDDKKPKGSGY 82

## RESULT 6

E56556

fork head homolog (clone gXPD-4) - African clawed frog (fragment)  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 12-Sep-1997  
C:Accession: E56556  
R:Knöchel, S.; Lef, J.; Clement, J.; Klocke, B.; Hille, S.; Koster, M.; Knöchel, W.  
Mech. Dev. 38, 157-165, 1992  
A:Title: Activin A induced expression of a fork head related gene in posterior chorda  
A:Reference number: A56556; MUID:93041288  
A:Accession: E56556  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-111 <KNO>  
A:Note: sequence extracted from NCBI backbone (NCBIP:118181)  
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
F:10-101/Domain: fork head DNA-binding domain homology <FHD>

## Query Match

Best Local Similarity 6.3%; Score 35; DB 2; Length 111;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 NKOGWNSIRHNSLNECFVKVPRDDKKPKGSGY 152  
Db 50 NKOGWNSIRHNSLNECFVKVPRDDKKPKGSGY 84

## RESULT 7

559870  
 fork head domain protein crocodile - fruit fly (*Drosophila melanogaster*)  
 C:Species: *Drosophila melanogaster*  
 C>Date: 19-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 21-Jul-2000  
 C:Accession: S59870; A46178  
 R:Kaeber, K.H.; Lee, K.H.; Schlondorff, J.; Hiemisch, H.; Monaghan, A.P.; Schutz, G.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 7628-7631, 1993  
 A:Title: Six members of the mouse forkhead gene family are developmentally regulated.  
 A:Reference number: A47746; MUID:93361500  
 A:Accession: F47746  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-508 <KAE>  
 R:Hacker, U.; Grossniklaus, U.; Gehring, W.J.; Jackle, H.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 8754-8758, 1992  
 A:Title: Developmentally regulated *Drosophila* gene family encoding the fork head domain.  
 A:Reference number: A46178; MUID:92409595  
 A:Accession: A46178  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 55-182 <HAC>  
 A:Cross-references: GB:96440; NID:9157425; PIDN:AAFP02177.1; PID:96042185  
 A:Note: sequence extracted from NCBI backbone (NCBIF:114222)  
 C:Genetics:  
 A:Gene: *CROC*  
 A:Cross-references: FlyBase:FBgn0014143  
 C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 F:70-161/Domain: fork head DNA-binding domain homology <FHD>

Query Match 4.5%; Score 25; DB 2; Length 508;  
 Best Local Similarity 100.0%; Pred. No. 3e-13; Indels 0; Gaps 0;  
 Matches 25; Conservative 0; Mismatches 0;  
 Oy 115 YRDNRKQGMNSIRHNLSECFVKV 139  
 ||||||||||||||||||||  
 Db 107 YRDNRKQGMNSIRHNLSECFVKV 131

## RESULT 8

447746  
 forkhead transcription factor fkh-3 - mouse (fragment)  
 C:Species: *Mus musculus* (house mouse)  
 C>Date: 03-Mar-1994 #sequence\_revision 05-Jan-1996 #text\_change 24-Sep-1999  
 C:Accession: C47746; S36072  
 R:Kaeber, K.H.; Lee, K.H.; Schlondorff, J.; Hiemisch, H.; Monaghan, A.P.; Schutz, G.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 7628-7631, 1993  
 A:Title: Six members of the mouse forkhead gene family are developmentally regulated.  
 A:Reference number: A47746; MUID:93361500  
 A:Accession: C47746  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-111 <KAE>  
 A:Cross-references: EMBL:X71944; NID:9311740; PIDN:CAAS0743.1; PID:9311741  
 A:Note: submitted to the EMBL Data Library, May 1993  
 A:Note: sequence extracted from NCBI backbone (NCBIF:136532)  
 C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 C:Keywords: DNA binding; transcription factor  
 F:10-101/Domain: fork head DNA-binding domain homology <FHD>

Query Match 4.2%; Score 23; DB 2; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-12; Indels 0; Gaps 0;  
 Matches 23; Conservative 0; Mismatches 0;  
 Oy 121 GMONSIRHNLSECFVKVPRDD 143  
 ||||||||||||||||||||  
 Db 53 GMONSIRHNLSECFVKVPRDD 75

## RESULT 9

447746  
 forkhead transcription factor fkh-6 - mouse (fragment)  
 C:Species: *Mus musculus* (house mouse)  
 C>Date: 03-Mar-1994 #sequence\_revision 05-Jan-1996 #text\_change 24-Sep-1999  
 C:Accession: F47746; S36075  
 R:Kaeber, K.H.; Lee, K.H.; Schlondorff, J.; Hiemisch, H.; Monaghan, A.P.; Schutz, G.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 7628-7631, 1993  
 A:Title: Six members of the mouse forkhead gene family are developmentally regulated.  
 A:Reference number: A47746; MUID:93361500  
 A:Accession: F47746  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-111 <KAE>  
 A:Cross-references: EMBL:X71944; NID:9311746; PIDN:CAAS0746.1; PID:9311747  
 A:Note: submitted to the EMBL Data Library, May 1993  
 A:Note: sequence extracted from NCBI backbone (NCBIF:136535)  
 C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 C:Keywords: DNA binding; transcription factor  
 F:10-101/Domain: fork head DNA-binding domain homology <FHD>

Query Match 4.0%; Score 22; DB 2; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 3e-11; Indels 0; Gaps 0;  
 Matches 22; Conservative 0; Mismatches 0;  
 Oy 120 GGMNSIRHNLSECFVKVPR 141  
 ||||||||||||||||||||  
 Db 52 GGMNSIRHNLSECFVKVPR 73

## RESULT 10

T16880  
 hypothetical protein T14G12.4 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jan-2000  
 C:Accession: T16880  
 R:Willcox, L.  
 submitted to the EMBL Data Library, November 1995  
 A:Description: The sequence of *C. elegans* cosmid T14G12.  
 A:Reference number: Z18596  
 A:Accession: T16880  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-270 <NTL>  
 A:Cross-references: EMBL:U41268; NID:91086843; PID:91086847; PIDN:AA82436.1; CESP:T1  
 C:Genetics:  
 A:Gene: *CESP:T14G12.4*  
 A:Introns: 37/1; 72/3; 164/1  
 C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
 F:93-185/Domain: fork head DNA-binding domain homology <FHD>

Query Match 3.8%; Score 21; DB 2; Length 270;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-10; Indels 0; Gaps 0;  
 Matches 21; Conservative 0; Mismatches 0;  
 Oy 113 PFYRDKQGMNSIRHNL 133  
 ||||||||||||||||||||  
 Db 128 PFYRDKQGMNSIRHNL 148

## RESULT 11

S23055  
 SLP2 protein - fruit fly (*Drosophila melanogaster*)  
 C:Species: *Drosophila melanogaster*  
 C>Date: 19-Feb-1994 #sequence\_revision 01-Sep-1995 #text\_change 24-Sep-1999  
 C:Accession: S23055  
 R:Grossniklaus, U.; Pearson, R.K.; Gehring, W.J.  
 Genes Dev. 6, 1030-1051, 1992  
 A:Title: The *Drosophila* sloppy paired locus encodes two proteins involved in segmenta  
 A:Reference number: S23053; MUID:92275347  
 A:Accession: S23055  
 A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-445 <GRO>  
A:Cross-references: EMBL:X66098; NID:98620; PIDN:CAA46892.1; PTD:98621  
C:Genetics:  
A:Gene: FLYBase:slp2  
A:Cross-references: FLYBase:FBgn0004567  
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
C:Keywords: DNA binding; transcription regulation  
F:180-271/Domain: fork head DNA-binding domain homology <FHD>

Query Match  
Best Local Similarity 3.4%; Score 19; DB 2; Length 445;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 YRDNKQGMNSIRHNSLN 133  
Db 217 YRDNKQGMNSIRHNSLN 235  
|||||

## RESULT 12

S1p2 protein - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 19-Feb-1994 #sequence\_revision 01-Sep-1995 #text\_change 24-Sep-1999  
C:Accession: S23056  
R:Grossniklaus, U.; Pearson, R.K.; Gehring, W.J.  
A:Title: The Drosophila sloppy paired locus encodes two proteins involved in segmental  
A:Reference number: S23053; MUID:92275347  
A:Accession: S23056  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-445 <GRO>  
A:Cross-references: EMBL:X66097; NID:98622; PIDN:CAA46891.1; PTD:98623  
C:Genetics:  
A:Gene: FLYBase:slp2  
A:Cross-references: FLYBase:FBgn0004567  
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
F:180-271/Domain: fork head DNA-binding domain homology <FHD>

Query Match  
Best Local Similarity 3.4%; Score 19; DB 2; Length 445;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 YRDNKQGMNSIRHNSLN 133  
Db 217 YRDNKQGMNSIRHNSLN 235  
|||||

## RESULT 13

FRBAC-7 protein - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 24-Sep-1999  
C:Accession: S51630  
R:Pierrou, S.; Helioglyst, M.; Sammelsson, L.; Enribeck, S.; Carlsson, P.  
A:Title: Cloning and characterization of seven human forkhead proteins: binding site spe  
A:Reference number: S51630; MUID:95045392  
A:Accession: S51630  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-106 <PIE>  
A:Cross-references: EMBL:U13225; NID:9563169; PIDN:AAA92042.1; PTD:9563170  
C:Genetics: GDB:FKHL1; FRBAC7  
A:Cross-references: GDB:450231  
A:Map position: 5q34-5q34  
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
F:6-97/Domain: fork head DNA-binding domain homology <FHD>

Query Match  
Best Local Similarity 2.9%; Score 16; DB 2; Length 106;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 TINGYOFIMDRFPY 115  
Db 28 TINGYOFIMDRFPY 43  
|||||

## RESULT 14

transcription factor HFR2 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 12-Sep-1997  
C:Accession: B54743  
R:Murphy, D.B.; Wiese, S.; Burfeind, P.; Schmundt, D.; Mattei, M.G.; Schulz-Schaeffer  
A:Title: Human brain factor 1, a new member of the fork head gene family.  
A:Reference number: A54743; MUID:95048332  
A:Accession: B54743  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-109 <MUR>  
A:Cross-references: GB:X74143  
C:Genetics:  
A:Gene: GDB:FKHL2; HFR-2; HFR2  
A:Cross-references: GDB:375747; OMIM:600779  
A:Map position: 14q11-14q13  
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
F:8-99/Domain: fork head DNA-binding domain homology <FHD>

Query Match  
Best Local Similarity 2.9%; Score 16; DB 2; Length 109;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKQGMNSIRHNSLN 133  
Db 48 NKQGMNSIRHNSLN 63  
|||||

## RESULT 15

Probable transcription factor fork head domain 2 (FD2) - fruit fly (Drosophila melano  
C:Species: Drosophila melanogaster  
C:Date: 22-Sep-1993 #sequence\_revision 25-Apr-1997 #text\_change 24-Sep-1999  
C:Accession: B46178  
R:Haeker, U.; Grossniklaus, U.; Gehring, W.J.; Jackle, H.  
A:Title: Developmentally regulated Drosophila gene family encoding the fork head doma  
A:Reference number: A46178; MUID:92409595  
A:Accession: B46178  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-134 <HAC>  
A:Cross-references: GB:M6441; NID:9157426; PIDN:AAA28533.1; PTD:9157427  
A:Note: Sequence extracted from NCBI backbone (NCBIP:114224)  
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology  
F:16-113/Domain: fork head DNA-binding domain homology <FHD>

Query Match  
Best Local Similarity 2.9%; Score 16; DB 2; Length 134;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKQGMNSIRHNSLN 133  
Db 56 NKQGMNSIRHNSLN 71  
|||||



Wed Feb 7 14:05:59 2001

Search completed: February 6, 2001, 23:15:18  
Job time: 8142 sec

us-09-292-862-2.rpt



Wed Feb 7 14:06:00 2001

US-09-292-862-2.rsp

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 6, 2001, 22:52:59 ; Search time 49.7 Seconds  
(without alignments)  
359.329 Million cell updates/sec

Title: US-09-292-862-2

Sequence: 553  
1 MQARYSVSPNSLGVVPLG.....PSSQSLVRTSGAFVYDCKSF 553

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 88757 seqs, 32294092 residues

Word size: 0

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Swissprot\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	250	45.2	553	1	FXCL_HUMAN
2	120	21.7	553	1	FXCL_MOUSE
3	61	11.0	494	1	FXC2_MOUSE
4	61	11.0	501	1	FXC2_HUMAN
5	59	10.7	101	1	FXC2_RAT
6	25	4.5	508	1	CROC_DROME
7	23	4.2	111	1	FXC2_HUMAN
8	23	4.2	111	1	FXC2_MOUSE
9	22	4.0	337	1	FXC2_MOUSE
10	19	3.4	106	1	FXC2_MOUSE
11	16	2.9	134	1	FXC2_MOUSE
12	16	2.9	134	1	FXC2_MOUSE
13	16	2.9	451	1	FXC2_MOUSE
14	16	2.9	451	1	FXC2_MOUSE
15	16	2.9	451	1	FXC2_MOUSE
16	16	2.9	451	1	FXC2_MOUSE
17	16	2.9	451	1	FXC2_MOUSE
18	16	2.9	451	1	FXC2_MOUSE
19	16	2.9	451	1	FXC2_MOUSE
20	15	2.7	101	1	FXC2_MOUSE
21	15	2.7	101	1	FXC2_MOUSE
22	15	2.7	101	1	FXC2_MOUSE
23	15	2.7	101	1	FXC2_MOUSE
24	15	2.7	101	1	FXC2_MOUSE
25	15	2.7	101	1	FXC2_MOUSE
26	15	2.7	101	1	FXC2_MOUSE
27	15	2.7	101	1	FXC2_MOUSE
28	15	2.7	101	1	FXC2_MOUSE
29	14	2.5	444	1	FXC2_MOUSE
30	14	2.5	444	1	FXC2_MOUSE
31	13	2.4	101	1	FXC2_MOUSE
32	13	2.4	101	1	FXC2_MOUSE
33	13	2.4	101	1	FXC2_MOUSE

Result ID	Score	Query Match	Length	DB ID	Description
34	13	2.4	443	1	OC3N_HUMAN
35	13	2.4	445	1	OC3N_MOUSE
36	13	2.4	543	1	YF91_MYCTU
37	13	2.4	617	1	FXK1_MOUSE
38	13	2.4	655	1	IFL1_HUMAN
39	13	2.4	862	1	FXH2_YEAST
40	12	2.2	106	1	FXD4_HUMAN
41	12	2.2	421	1	FXD4_HUMAN
42	12	2.2	421	1	FXD4_HUMAN
43	12	2.2	421	1	FXD4_HUMAN
44	12	2.2	421	1	FXD4_HUMAN
45	12	2.2	564	1	HCM1_YEAST

# ALIGNMENTS

Result ID	Score	Query Match	Length	DB ID	Description
1	250	45.2	553	1	FXCL_HUMAN
2	120	21.7	553	1	FXCL_MOUSE
3	61	11.0	494	1	FXC2_MOUSE
4	61	11.0	501	1	FXC2_HUMAN
5	59	10.7	101	1	FXC2_RAT
6	25	4.5	508	1	CROC_DROME
7	23	4.2	111	1	FXC2_HUMAN
8	23	4.2	111	1	FXC2_MOUSE
9	22	4.0	337	1	FXC2_MOUSE
10	19	3.4	106	1	FXC2_MOUSE
11	16	2.9	134	1	FXC2_MOUSE
12	16	2.9	134	1	FXC2_MOUSE
13	16	2.9	451	1	FXC2_MOUSE
14	16	2.9	451	1	FXC2_MOUSE
15	16	2.9	451	1	FXC2_MOUSE
16	16	2.9	451	1	FXC2_MOUSE
17	16	2.9	451	1	FXC2_MOUSE
18	16	2.9	451	1	FXC2_MOUSE
19	16	2.9	451	1	FXC2_MOUSE
20	15	2.7	101	1	FXC2_MOUSE
21	15	2.7	101	1	FXC2_MOUSE
22	15	2.7	101	1	FXC2_MOUSE
23	15	2.7	101	1	FXC2_MOUSE
24	15	2.7	101	1	FXC2_MOUSE
25	15	2.7	101	1	FXC2_MOUSE
26	15	2.7	101	1	FXC2_MOUSE
27	15	2.7	101	1	FXC2_MOUSE
28	15	2.7	101	1	FXC2_MOUSE
29	14	2.5	444	1	FXC2_MOUSE
30	14	2.5	444	1	FXC2_MOUSE
31	13	2.4	101	1	FXC2_MOUSE
32	13	2.4	101	1	FXC2_MOUSE
33	13	2.4	101	1	FXC2_MOUSE

entires requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AF048693; AAC18081.1;  
 CC EMBL: U13221; AAA92038.1;  
 DR EMBL: AF078096; AAC72915.1;  
 DR MIM: 601090;  
 DR MIM: 601631;  
 DR INTERPRO: IPR001766;  
 DR INTERPRO: IPR002952;  
 DR INTERPRO: IPR002965;  
 DR PFM: PF00250; FORK\_HEAD\_1;  
 DR PRINTS: PR00053; FORKHEAD;  
 DR PRINTS: PR01217; PRICHEXTENSIN;  
 DR PRINTS: PR01228; EGGSHELL;  
 DR PROSITE: PS00657; FORK\_HEAD\_1;  
 DR PROSITE: PS00658; FORK\_HEAD\_2;  
 DR PROSITE: PS50039; FORK\_HEAD\_3;  
 DR DNA-binding; Nuclear protein; Transcription regulation;  
 FW Disease mutation;  
 FT DNA\_BIND 77 168  
 FT DOMAIN 28 33  
 FT DOMAIN 169 173  
 FT DOMAIN 194 197  
 FT DOMAIN 262 272  
 FT DOMAIN 292 297  
 FT DOMAIN 375 382  
 FT DOMAIN 438 445  
 FT DOMAIN 447 456  
 FT DOMAIN 486 495  
 FT VARIANT 82 85  
 FT VARIANT 87 87  
 FT VARIANT 112 112  
 FT VARIANT 126 126  
 FT VARIANT 131 131  
 FT CONFLICT 180 180  
 FT CONFLICT 199 202  
 FT CONFLICT 426 426  
 FT SEQUENCE 553 AA; 56787 MW; DBC81B9430359E CRC64;  
 Query Match 45.2%; Score 250; DB 1; Length 553;  
 Best Local Similarity 99.7%; Pred No. 1.6e-192;  
 Matches 350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CC EMBL: AF048693; AAC18081.1;  
 CC EMBL: U13221; AAA92038.1;  
 DR EMBL: AF078096; AAC72915.1;  
 DR MIM: 601090;  
 DR MIM: 601631;  
 DR INTERPRO: IPR001766;  
 DR INTERPRO: IPR002952;  
 DR INTERPRO: IPR002965;  
 DR PFM: PF00250; FORK\_HEAD\_1;  
 DR PRINTS: PR00053; FORKHEAD;  
 DR PRINTS: PR01217; PRICHEXTENSIN;  
 DR PRINTS: PR01228; EGGSHELL;  
 DR PROSITE: PS00657; FORK\_HEAD\_1;  
 DR PROSITE: PS00658; FORK\_HEAD\_2;  
 DR PROSITE: PS50039; FORK\_HEAD\_3;  
 DR DNA-binding; Nuclear protein; Transcription regulation;  
 FW Disease mutation;  
 FT DNA\_BIND 77 168  
 FT DOMAIN 28 33  
 FT DOMAIN 169 173  
 FT DOMAIN 194 197  
 FT DOMAIN 262 272  
 FT DOMAIN 292 297  
 FT DOMAIN 375 382  
 FT DOMAIN 438 445  
 FT DOMAIN 447 456  
 FT DOMAIN 486 495  
 FT VARIANT 82 85  
 FT VARIANT 87 87  
 FT VARIANT 112 112  
 FT VARIANT 126 126  
 FT VARIANT 131 131  
 FT CONFLICT 180 180  
 FT CONFLICT 199 202  
 FT CONFLICT 426 426  
 FT SEQUENCE 553 AA; 56787 MW; DBC81B9430359E CRC64;  
 Query Match 45.2%; Score 250; DB 1; Length 553;  
 Best Local Similarity 99.7%; Pred No. 1.6e-192;  
 Matches 350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 2  
 EXCL\_MOUSE STANDARD: PRT; 553 AA.  
 AC 061572; 061582; 088409;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE FORKHEAD BOX PROTEIN C1 (FORKHEAD-RELATED PROTEIN FKHL7) (FORKHEAD-RELATED TRANSCRIPTION FACTOR 3) (FRC-3) (TRANSCRIPTION FACTOR FKHL-1) (MESODERM/MESENCHYME FORKHEAD 1) (MF-1).  
 GN FOXCL OR FKHL7 OR FRCAC3 OR FKHL OR MF1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=98297351; PubMed=9635428;  
 RA Kume T., Deng K.Y., Winfrey V., Gould D.B., Walter M.A., Hogan B.L.M.;  
 RT "The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus";  
 RL Cell 93:985-996(1998).  
 RN [2]  
 RP SEQUENCE OF 69-179 FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=93361500; PubMed=7689224;  
 RX Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H., Monaghan A.P., Schuetz G.;  
 RA "Six members of the mouse forkhead gene family are developmentally regulated";  
 RT Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).  
 RN [3]  
 RP SEQUENCE OF 71-187 FROM N.A.  
 RX MEDLINE=93387221; PubMed=8375339;  
 RX Sasaki H., Hogan B.L.;  
 RT "Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo";  
 RL Development 118:47-59(1993).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN MANY EMBRYONIC TISSUES, INCLUDING PRECHONDROGENIC MESENCHYME, PERICULAR MESENCHYME, MENINGES, ENDOTHELIAL CELLS, AND KIDNEY.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

FT DOMAIN 444 451 POLY-SER.  
 FT DOMAIN 453 456 POLY-GLY.  
 FT DOMAIN 486 496 POLY-ALA.  
 FT CONFLICT 180 187 VKDKEEGK -> KKEITFIG (IN REF. 3).  
 SO SEQUENCE 553 AA: 56953 MW: 3CDD12F69CA4F217 CRC64;

Query Match 21.7%: Score 120; DB 1; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-88;  
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 MARAYGPTPOPKDMVKPPYSYIALITMAIONAPDKKITTITNGIYOFIMDFPPYRDNK 119  
 DB 60 MARAYGPTPOPKDMVKPPYSYIALITMAIONAPDKKITTITNGIYOFIMDFPPYRDNK 119  
 OY 120 OGOMNSIRHNLNLNCEFCYKVPDRDKKPGKGSYWTLPDPSYNMFENGSLFRRRRRKKKDA 179  
 DB 120 OGOMNSIRHNLNLNCEFCYKVPDRDKKPGKGSYWTLPDPSYNMFENGSLFRRRRRKKKDA 179

RESULT 3  
 FXC2\_MOUSE STANDARD: PRT: 494 AA.  
 ID FXC2\_MOUSE 061850; Q63869; P97948;  
 AC 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE FORKHEAD BOX PROTEIN C2 (FORKHEAD-RELATED PROTEIN FKHL14) (MESENCHYME  
 DE FORK HEAD PROTEIN 1) (MFH-1 PROTEIN) (TRANSCRIPTION FACTOR FKHL-14).  
 GN FOXC2 OR FKHL14 OR MFH1 OR FKHL4.  
 OS Mus musculus (Mouse);  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RC STRAIN=C57BL/6;  
 RA Kaestner K.H., Blackmann S., Monaghan A.P., Schlondorff J.,  
 RA Muncheva A., Lichter P., Schutz G.;  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129;  
 RX MEDLINE=97312712; PubMed=9169153;  
 RX Miura N., Iida K., Kakimura H., Yang X.-L., Sugiyama T.;  
 RT "Isolation of the mouse (MFH-1) and human (FKHL14) mesenchyme fork  
 RT head-1 genes reveals conservation of their gene and protein  
 RT structures";  
 RL Genomics 41:489-492(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ICR;  
 RA Miura N.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 34-494 FROM N.A.  
 RX MEDLINE=93314779; PubMed=8525367;  
 RX Miura N., Watanaka A., Tohyama M., Tanaka K.;  
 RT "MFH-1, a new member of the fork head domain family, is expressed in  
 RT developing mesenchyme";  
 RL FEBS Lett. 326:171-176(1993).  
 CC -1- FUNCTION: MIGHT BE INVOLVED IN THE FORMATION OF SPECIAL  
 CC MESENCHYMAL TISSUES.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED TEMPORALLY IN DEVELOPING EMBRYOS,  
 CC FIRST IN THE NON-NOTOCHORDAL MESODERM AND LATER IN AREAS OF  
 CC MESENCHYMAL CONDENSATION IN THE TRUNK, HEAD, AND LIMBS.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X92499; CA63244.1; -  
 CC DR EMBL: Y08222; CA69399.1; -  
 CC DR EMBL: X74040; CA652192.1; -  
 CC DR EMBL: S63607; AAB27463.1; -  
 CC DR TRANSFAC: T02445; -  
 CC DR MGD: MGI:1347481; FOXC2.  
 CC DR INTERPRO: IPR001766; -  
 CC DR PFAM: PF00250; Fork\_head; 1.  
 CC PRINTS: PR00053; FORKHEAD.  
 CC DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 CC DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 CC DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 CC KW Transcription regulation; DNA-binding; Nuclear protein.  
 FT DNA BIND 70 161 FORK-HEAD.  
 FT DOMAIN 162 166 POLY-ARG.  
 FT DOMAIN 366 395 HIS-RICH.  
 FT DOMAIN 396 415 ALA/PRO-RICH.  
 FT CONFLICT 12 13 AL -> V (IN REF. 1).  
 SO SEQUENCE 494 AA: 52874 MW: 232C0D6FA64320A6 CRC64;

Query Match 11.0%: Score 61; DB 1; Length 494;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-41;  
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 NKOGMNSIRHNLNLNCEFCYKVPDRDKKPGKGSYWTLPDPSYNMFENGSLFRRRRRKKK 177  
 DB 111 NKOGMNSIRHNLNLNCEFCYKVPDRDKKPGKGSYWTLPDPSYNMFENGSLFRRRRRKKK 170  
 OY 178 D 178  
 DB 171 D 171

RESULT 4  
 FXC2\_HUMAN STANDARD: PRT: 501 AA.  
 ID FXC2\_HUMAN 099958;  
 AC 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE FORKHEAD BOX PROTEIN C2 (FORKHEAD-RELATED PROTEIN FKHL14) (MESENCHYME  
 DE FORK HEAD PROTEIN 1) (MFH-1 PROTEIN) (TRANSCRIPTION FACTOR FKHL-14).  
 GN FOXC2 OR FKHL14 OR MFH1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 NC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 RC [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97312712; PubMed=9169153;  
 RX Miura N., Iida K., Kakimura H., Yang X.-L., Sugiyama T.;  
 RT "Isolation of the mouse (MFH-1) and human (FKHL14) mesenchyme fork  
 RT head-1 genes reveals conservation of their gene and protein  
 RT structures";  
 RL Genomics 41:489-492(1997).  
 CC -1- FUNCTION: MIGHT BE INVOLVED IN THE FORMATION OF SPECIAL  
 CC MESENCHYMAL TISSUES (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; "The genome sequence of *Drosophila melanogaster*."; Science 287:2185-2195(2000).

[3]  
 RN SEQUENCE OF 55-182 FROM N.A.  
 RP MEDLINE=92409595; PubMed=1356269;  
 RX Haacker U., Grossniklaus U., Gehring W.J., Jaechle H.;  
 RA "Developmentally regulated Drosophila gene family encoding the fork  
 RT head domain";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:8754-8758(1992).  
 CC -1- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OF HEAD STRUCTURES.  
 CC REQUIRED TO FUNCTION AS AN EARLY PATTERNING GENE IN THE ANTERIOR OF A  
 CC MOST BLASTODERM HEAD SEGMENT ANLAGE AND FOR THE ESTABLISHMENT OF A  
 CC SPECIFIC HEAD SKELETAL STRUCTURE THAT DERIVES FROM THE NON-  
 CC ADJACENT INTERCALARY SEGMENT AT A LATER STAGE OF EMBRYOGENESIS.  
 CC BINDS THE CONSENSUS DNA SEQUENCE 5'-(AG)TAAATTC(A)-3'.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED INITIALLY IN BOTH THE ANTERIOR AND  
 CC POSTERIOR REGIONS OF THE EMBRYO.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: S80254; AAB35643.1; -  
 CC EMBL: AF003594; AAF51727.1; -  
 CC EMBL: M96440; AAF02177.1; -  
 CC PIR: A46178; A46178.  
 CC TRANSFAC: T02291; -  
 CC FLYBASE: FBgn0014143; CROC.  
 CC INTERPRO: IPR001766; -  
 CC PIR: PF00250; FORK\_HEAD\_1; 1.  
 CC PRINTS: PS00657; FORK\_HEAD\_2; 1.  
 CC PROSITE: PS00658; FORK\_HEAD\_3; 1.  
 CC PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 CC DR DNA-binding: Developmental protein; Nuclear protein;  
 CC KW Transcription regulation.  
 CC FT DOMAIN 34 40 POLY-ALA.  
 CC FT DNA\_BIND 69 160 FORK-HEAD.  
 CC FT DOMAIN 161 165 POLY-ARG.  
 CC FT DOMAIN 301 304 POLY-ALA.  
 CC FT DOMAIN 377 380 POLY-ASN.  
 CC FT DOMAIN 389 403 POLY-GLY.  
 CC FT DOMAIN 452 461 POLY-ALA.  
 CC FT DOMAIN 466 473 POLY-HIS.  
 CC FT VARIANT 122 122 L -> F (IN ALLELE CROC-75-3).  
 CC FT VARIANT 453 453 A -> V (IN ALLELE CROC-75-3).  
 CC FT SEQUENCE 508 AA; 54516 MW; 2EFEDID8F63016D6 CRC64;  
 SQ  
 Query Match 4.5%; Score 25; DB 1; Length 508;  
 Best Local Similarity 100.0%; Pred. No. 1,3e-12;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=97468148; PubMed=9325056;  
 RX Cedergren A., Betz R., Lagercrantz S., Larsson C., Hultander M.,  
 RA Carlsson P., Eneback S.;  
 RA "Chromosome localization, sequence analysis, and expression pattern  
 RT identify FKL 18 as a novel human forkhead gene";  
 RL Genomics 44:344-346(1997).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF042831; AAC15420.1; -  
 CC MIM: 602939; -  
 CC DR INTERPRO: IPR001766; -  
 CC DR PIR: PF00250; FORK\_HEAD\_1; 1.  
 CC PROSITE: PS00657; FORK\_HEAD\_2; 1.  
 CC PROSITE: PS00658; FORK\_HEAD\_3; 1.  
 CC PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 CC DR Transcription regulation; DNA-binding; Nuclear protein.  
 CC KW Transcription regulation; DNA-binding; Nuclear protein.  
 CC FT NON\_TER 1 100 FORK-HEAD.  
 CC FT DNA\_BIND 9 100 FORK-HEAD.  
 CC FT NON\_TER 111 111  
 CC FT SEQUENCE 111 AA; 12904 MW; BCEALID6AC84B64F3 CRC64;  
 SQ

Query Match 4.2%; Score 23; DB 1; Length 111;  
 Best Local Similarity 100.0%; Pred. No. 1,5e-11;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 121 GWNSIRHNLSECEFKVPRD 143  
 DB 53 GWNSIRHNLSECEFKVPRD 75  
 RESULT 8  
 FREA\_MOUSE STANDARD; PRT: 111 AA.  
 ID FREA\_MOUSE  
 AC 061574;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE FORKHEAD-RELATED TRANSCRIPTION FACTOR 10 (FREC-10) (TRANSCRIPTION  
 DE FACTOR FKH-3) (FRAGMENT)  
 GN FKH18 OR FREAC10 OR FKH3.  
 GN Mus musculus (Mouse).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Eukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=129;  
 RX MEDLINE=93361500; PubMed=7689224;  
 RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H.,  
 RA Monaghan A.P., Schuetz G.;  
 RA "Six members of the mouse forkhead gene family are developmentally  
 RT regulated";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way



modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: X71941: CAA50743.1; -  
 TRANSFAC: T02440; -  
 DR MGI:95546; FKH3.  
 DR INTERPRO: IPR001766; -  
 DR PFM: PF00250; Fork\_head; 1.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 KW Transcription regulation; DNA-binding; Nuclear protein.  
 FT DNA\_BIND 1 100 - FORK-HEAD.  
 FT NON\_TER 111 111  
 SO SEQUENCE 111 AA; 12905 MW; 22492EC48BC05CE1 CRC64;

Query Match  
 Best Local Similarity 4.2%; Score 23; DB 1; Length 111;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 121 QGMONSTRHNLINCECFVYPRD 143  
 DB 53 QGMONSTRHNLINCECFVYPRD 75

RESULT 9  
 ID FX1L\_MOUSE STANDARD; PRT; 337 AA.  
 AC 064731;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE FORKHEAD PROTEIN L1 (FORKHEAD-RELATED PROTEIN FKHL1) (TRANSCRIPTION FACTOR FKHL1 OR FKHL1 OR FKHL1)  
 GN FOXL1 OR FKHL1 OR FKHL1  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-129;  
 RX MEDLINE=96322289; PubMed=8674414;  
 RA Kestner K.H., Bleckmann S.C., Monaghan A.P., Schlondorff J., Mincheva A., Licher P., Schuetz G.;  
 RT "Clustered arrangement of winged helix genes fkh-6 and MFR-1: possible implications for mesoderm development.";  
 RL Development 122:1751-1758(1996).  
 RN [2]  
 RP SEQUENCE OF 40-150 FROM N.A.  
 RC STRAIN-129;  
 RX MEDLINE=93361500; PubMed=7689224;  
 RA Kestner K.H., Lee K.H., Schloendorff J., Hiemisch H., Monaghan A.P., Schuetz G.;  
 RT "Six members of the mouse forkhead gene family are developmentally regulated.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: X92498; CAA63243.1; -  
 EMBL: X71944; CAA50746.1; -  
 TRANSFAC: T02417; -

Query Match  
 Best Local Similarity 4.0%; Score 22; DB 1; Length 337;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 120 QGMONSTRHNLINCECFVYPR 141  
 DB 91 QGMONSTRHNLINCECFVYPR 112

RESULT 10  
 ID SLP2\_DROME STANDARD; PRT; 445 AA.  
 AC P32031;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 DE FORK HEAD DOMAIN TRANSCRIPTION FACTOR SLP2 (SLOPPY PAIRED LOCUS PROTEIN 2).  
 GN SLP2.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTON-S, AND OREGON-R;  
 RX MEDLINE=92275347; PubMed=1317319;  
 RA Grossniklaus U., Pearson R.K., Gehring M.J.;  
 RT "The Drosophila sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors.";  
 RL Genes Dev. 6:1030-1051(1992).  
 CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN SEGMENTATION. MAY OF SLP ACTIVITY SEEM TO BE REQUIRED IN DIFFERENT LEVELS OF SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE POSTERIOR HALF OF EACH PARASEGMENT JUST ANTERIOR TO THE PARASEGMENTAL BOUNDARY.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION AT 3-6 HRS OF EMBRYOGENESIS.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Transcription regulation; Segmentation polarity protein.  
 KW DNA\_BIND 179 270 FORK-HEAD.  
 FT DNAS 94 104 ASN-RICH.  
 FT DOMAIN 362 362 R -> P (IN OREGON-R).  
 FT VARIANT 378 378 S -> N (IN OREGON-R).  
 FT VARIANT 387 387 O -> R (IN OREGON-R).  
 FT VARIANT 387 387 H -> Q (IN OREGON-R).  
 FT VARIANT 410 410 T -> I (IN OREGON-R).  
 FT VARIANT 416 416  
 SO SPOUNCE 445 AA; 50648 MW; 2CD253B7135B5F8D CRC64;

Query Match 3.4%; Score 19; DB 1; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-08; Indels 0; Gaps 0;  
 Matches 19; Conservative 0; Mismatches 0;

OY 115 YRDNGKQMSIRHNLSN 133  
 DB 217 YRDNGKQMSIRHNLSN 235

RESULT 11  
 ID EXL1\_HUMAN STANDARD; PRT; 106 AA.  
 AC 012952;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE FORKHEAD PROTEIN L1 (FORKHEAD-RELATED PROTEIN FKHL1) (FORKHEAD-RELATED TRANSCRIPTION FACTOR 7) (FREAC-7) (FRAGMENT).  
 DE RELATED TRANSCRIPTION FACTOR 7 (FREAC-7) (FRAGMENT).  
 GN FOXL1 OR FKHL1 OR FREAC7.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP MEDLINE=95045392; PubMed=7957066;  
 RA Pierson S., Hellqvist M., Samuelsson L., Emerbaeck S., Carlsson P.;  
 RT Cloning and characterization of seven human forkhead proteins;  
 RT binding site specificity and DNA bending.";  
 RT EMBL J. 13:5002-5012(1994).  
 RL EMBL J. 13:5002-5012(1994).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; U13225; AAA92042.1;  
 DR EMBL; 603252;  
 DR INTERPRO: IPR001766;  
 DR PFM: PF00250; Fork\_head\_1; 1.  
 DR PROSITE: PS00657; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_3; 1.  
 DR PROSITE: PS00659; FORK\_HEAD\_4; 1.  
 DR DNA-binding; Nuclear protein; Transcription regulation.  
 KW DNA-binding; Nuclear protein; Transcription regulation.  
 FT DNA\_BIND 5 96 FORK-HEAD.  
 FT NON\_TER 106 106  
 FT NON\_TER 106 106  
 SO SEQUENCE 106 AA; 12499 MW; 22136293EBB2BF17 CRC64;

Query Match 2.9%; Score 16; DB 1; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 5.8e-06; Indels 0; Gaps 0;  
 Matches 16; Conservative 0; Mismatches 0;

OY 100 TLNGIYQIMDRPPY 115  
 DB 28 TLNGIYQIMDRPPY 43

RESULT 12  
 ID FD2\_DROME STANDARD; PRT; 134 AA.  
 AC 002360;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 DE FORK HEAD DOMAIN PROTEIN FD2 (FRAGMENT).  
 DE FD64 OR FD2.  
 GN Drosophila melanogaster (Fruit fly).  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 RN [1]  
 RP SPOUNCE FROM N.A.  
 RP MEDLINE=92409595; PubMed=1356269;  
 RA Haacker U., Grossniklaus U., Gehring W.J., Jaekle H.;  
 RT Developmentally regulated Drosophila gene family encoding the fork head domain.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 89:8754-8758(1992)  
 RL head domain.";  
 CC -1- FUNCTION: INVOLVED IN DEVELOPMENT DURING EMBRYOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; M96441; AAA28533.1;  
 DR EMBL; B46178; B46178.  
 DR EMBL; F860004895; fd64A.  
 DR INTERPRO: IPR001766;  
 DR PFM: PF00250; Fork\_head\_1; 1.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS00659; FORK\_HEAD\_3; 1.  
 DR DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation.  
 FT DNA\_BIND 1 1 FORK-HEAD.  
 FT NON\_TER 134 134  
 FT NON\_TER 134 134  
 SO SEQUENCE 134 AA; 15798 MW; 4DFD5DE0967E308 CRC64;

Query Match 2.9%; Score 16; DB 1; Length 134;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-06; Indels 0; Gaps 0;  
 Matches 16; Conservative 0; Mismatches 0;

OY 118 NKQGMNSIRHNLSN 133  
 DB 56 NKQGMNSIRHNLSN 71

RESULT 13  
 ID QIN\_AVIS3 STANDARD; PRT; 387 AA.  
 AC P56260;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE TRANSFORMING PROTEIN QIN (ONCOGENE QIN).  
 DE V-QIN.  
 GN Avian sarcoma virus (strain 31) (ASV31).  
 OS Avian sarcoma virus (strain 31) (ASV31).  
 OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.  
 RN [1]  
 RP SPOUNCE FROM N.A.  
 RP MEDLINE=93281605; PubMed=8099441;  
 RA Li J., Vogt P.K.;

RT "The retroviral oncogene gln belongs to the transcription factor  
 RT family that includes the homeotic gene fork head";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:4490-4494(1993).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-QIN  
 CC POLYPEPTIDE.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L10719; NOT ANNOTATED\_CDS.  
 DR INTERPRO: IPR001766; -  
 DR PRAM: PF00250; Fork\_head: 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS00039; FORK\_HEAD\_3; 1.  
 KW DNA-binding; Nuclear protein; Transforming protein; Oncogene.  
 FT DOMAIN 42 54 POLY-HIS.  
 FT DOMAIN 53 58 POLY-PRO.  
 FT DOMAIN 64 67 POLY-ALA.  
 FT DOMAIN 101 106 POLY-ALA.  
 FT DNA\_BIND 142 233 FORK-HEAD.  
 SQ SEQUENCE 387 AA; 42283 MW; FEA902F50FE42P9 CRC64;

Query Match  
 Best Local Similarity 2.9%; Score 16; DB 1; Length 387;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 NKOGWONSIRHNLSTN 133  
 DB 183 NKOGWONSIRHNLSTN 198  
 ||||||||||||||||  
 ||||||||||||||||

RESULT 14  
 EXGB\_CHICK STANDARD; PRT; 451 AA.  
 AC Q90964;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE FORKHEAD PROTEIN G1B (FORKHEAD-RELATED PROTEIN FKHL1) (TRANSCRIPTION  
 DE FACTOR BF-1) (BRAIN FACTOR 1) (BFL) (CBF-1) (PROTO-ONCOGENE C-QIN) (N-  
 DE 62-5) (CEO 3-1).  
 GN FOXG1B OR FKHL1 OR QIN.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95132616; PubMed=7831308;  
 RA Chang H.W., Li J., Kretschmar D., Vogt P.K.;  
 RA "Avian cellular homolog of the gln oncogene.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 92:447-451(1995).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WHITE LESHORN; TISSUE=RETINA;  
 RX MEDLINE=96338226; PubMed=8757134;  
 RA Yusa J., Hirano S., Yamagata M., Noda M.;  
 RT "Visual projection map specified by topographic expression of  
 RT transcription factors in the retina.";  
 RL Nature 382:632-635(1996).  
 CC -1- FUNCTION: MAY DETERMINE THE NASOTEMPORAL AXIS OF THE RETINA, AND  
 CC CONSEQUENTLY SPECIFY THE TOPOGRAPHICAL PROJECTION OF THE RETINAL  
 CC GANGLION-CELL AXONS TO THE TECTUM BY CONTROLLING EXPRESSION OF

CC  
 CC THEIR TARGET GENES.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: RETINA AND BRAIN.  
 CC -1- DEVELOPMENTAL STAGE: CAN BE DETECTED IN REGIONS INCLUDING  
 CC PRIMORDIAL RETINA AND NEUROEPITHELIUM BY EMBRYONIC DAY 2 (E2). AT  
 CC E3, EXPRESSED IN THE NASAL RETINA AND PIGMENT EPITHELIUM AS WELL  
 CC AS IN THE TECTUM, AND AT E7 IS EXPRESSED IN RETINAL  
 CC GANGLION CELLS. LEVELS BEGIN TO DECLINE FROM E4 AND ALMOST  
 CC DISAPPEAR BY E10.  
 CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U47275; AA08466.1; -  
 DR INTERPRO: IPR001766; -  
 DR PRAM: PF00250; Fork\_head: 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS00039; FORK\_HEAD\_3; 1.  
 KW Transcription regulation; DNA-binding; Nuclear protein;  
 KW Developmental protein; Proto-oncogene.  
 FT DOMAIN 45 54 POLY-HIS.  
 FT DOMAIN 55 58 POLY-PRO.  
 FT DOMAIN 64 67 POLY-ALA.  
 FT DOMAIN 103 106 POLY-ALA.  
 FT DNA\_BIND 142 233 FORK-HEAD.  
 SQ SEQUENCE 451 AA; 48856 MW; E9E5B407D321B50 CRC64;

Query Match  
 Best Local Similarity 2.9%; Score 16; DB 1; Length 451;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 NKOGWONSIRHNLSTN 133  
 DB 183 NKOGWONSIRHNLSTN 198  
 ||||||||||||||||  
 ||||||||||||||||

RESULT 15  
 EXGA\_HUMAN STANDARD; PRT; 469 AA.  
 AC P53316;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE FORKHEAD PROTEIN G1A (FORKHEAD-RELATED PROTEIN FKHL2) (TRANSCRIPTION  
 DE FACTOR BF-2) (BRAIN FACTOR 2) (BF2) (HFK2).  
 GN FOXG1A OR FKHL2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=95048332; PubMed=7959731;  
 RA Murphy D.B., Wiese S., Burfield P., Schmundt D., Mettel M.-G.,  
 RA Schulz-Schaeffer W., Thies U.;  
 RT "Human brain factor 1, a new member of the fork head gene family.";  
 RL Genomics 21:551-557(1994).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95322450; PubMed=7599184;  
 RA Wiese S., Murphy D.B., Schilling A., Burfield P., Schmundt D.,  
 RA Schuller V., Mettel M.-G., Thies U.;  
 RT "The genes for human brain factor 1 and 2, members of the fork head

```

RT gene family, are clustered on chromosome 14q."
RL Blochm. Biophys. Acta 1262:105-112(1995).
CC -I- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ESTABLISHMENT OF THE
CC REGIONAL SUBDIVISION OF THE DEVELOPING BRAIN AND IN THE
CC DEVELOPMENT OF THE TELECEPHALON. SEQUENCE-SPECIFIC DNA-BINDING
CC PROTEIN WITH A DISTINCT BINDING SPECIFICITY (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -I- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X78202; CAA55038.1; -
DR EMBL: X74143; CAA52240.1; -
DR TRANSFAC: T02292; -
DR MIM: 600779; -
DR INTERPRO: IPR001766; -
DR PFAM: PF00250; FORK_HEAD_1.
DR PRINTS: PR00053; FORKHEAD.
DR PROSITE: PS00657; FORK_HEAD_1; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS50039; FORK_HEAD_3; 1.
DR KW transcription regulation; DNA-binding; Nuclear protein;
KW Developmental protein.
KW DOMAIN 33 56 HIS-RICH.
FT DOMAIN 57 79 PRO-RICH.
FT DOMAIN 70 73 POLY-GLN.
FT DOMAIN 83 88 POLY-ARG.
FT DNA_BIND 161 252 FORK-HEAD.
SQ SEQUENCE 469 AA; 50539 MW; 943B8BDB90008EDC CRC64;

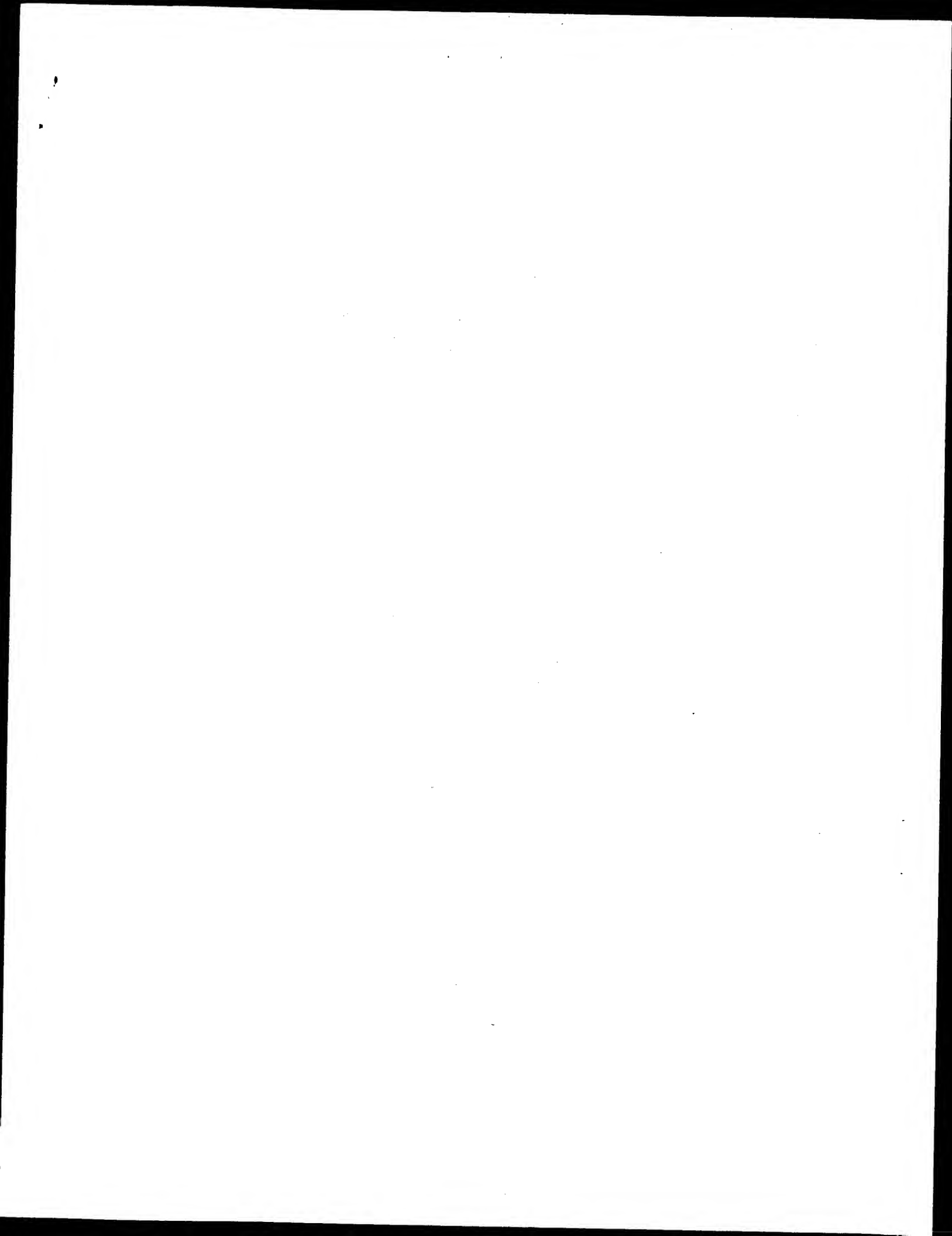
```

```

Query Match 2.9%; Score 16; DB 1; Length 469;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 118 NKQGMNSIRHNLSLN 133
Db 202 NKQGMNSIRHNLSLN 217

```

Search completed: February 7, 2001, 00:41:05  
 Job time: 6486 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: February 6, 2001, 20:25:51 ; Search time 51.67 Seconds  
(without alignments)  
365.960 Million cell updates/sec

Title: US-09-292-862-2  
 Perfect score: 553  
 Sequence: 1 MQARYSVSPNSLGVVPLG.....PSSQSLYRTSGAFVYDCKF 553

```

Scoring table:
  Oligo      Gapop  60.0 ,  Gapext  60.0
Searched:    268485 seqs, 34193795 residues

```

```

Word size :      0
Total number of hits satisfying chosen parameters: 268485

```

Minimum	DB seq	length:	0
Maximum	DB seq	length:	20000000000

Post-processing: Listing first 45 summaries

1:	/SID56/gcgdata1/geneseq/genesep/AI1980.DAT *
2:	/SID56/gcgdata1/geneseq/genesep/AI1980.DAT *
3:	/SID56/gcgdata1/geneseq/genesep/AI1982.DAT *
4:	/SID56/gcgdata1/geneseq/genesep/AI1983.DAT *
5:	/SID56/gcgdata1/geneseq/genesep/AI1984.DAT *
6:	/SID56/gcgdata1/geneseq/genesep/AI1985.DAT *
7:	/SID56/gcgdata1/geneseq/genesep/AI1986.DAT *
8:	/SID56/gcgdata1/geneseq/genesep/AI1987.DAT *
9:	/SID56/gcgdata1/geneseq/genesep/AI1988.DAT *
10:	/SID56/gcgdata1/geneseq/genesep/AI1989.DAT *
11:	/SID56/gcgdata1/geneseq/genesep/AI1990.DAT *
12:	/SID56/gcgdata1/geneseq/genesep/AI1991.DAT *
13:	/SID56/gcgdata1/geneseq/genesep/AI1992.DAT *
14:	/SID56/gcgdata1/geneseq/genesep/AI1993.DAT *
15:	/SID56/gcgdata1/geneseq/genesep/AI1994.DAT *
16:	/SID56/gcgdata1/geneseq/genesep/AI1995.DAT *
17:	/SID56/gcgdata1/geneseq/genesep/AI1996.DAT *
18:	/SID56/gcgdata1/geneseq/genesep/AI1997.DAT *
19:	/SID56/gcgdata1/geneseq/genesep/AI1998.DAT *
20:	/SID56/gcgdata1/geneseq/genesep/AI1999.DAT *
21:	/SID56/gcgdata1/geneseq/genesep/AI2000.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	553	100.0	53	20	Y41277	Human FHEA3 prote
2	250	45.2	53	20	Y55721	Human forkhead tra
3	223	40.3	53	20	Y43260	Human FKH17 protel
4	61	11.0	501	20	Y01097	FHEA11 protein.
5	48	8.7	106	20	Y55723	FKH17 forkhead dom
6	40	7.2	106	20	Y55724	FKH14 forkhead do
7	17	3.1	106	20	Y55725	FKH18 forkhead do
8	16	2.9	105	20	Y55735	FKH11 forkhead do
9	16	2.9	106	20	Y55726	FKH14 forkhead dom
10	16	2.9	106	20	Y55734	FKH11 forkhead do
11	16	2.9	480	14	R44551	Brain factor-1, F
12	13	2.4	76	20	Y55737	FKH10 forkhead do

13	13	2.4	106	20	Y55729	FKHL17 forkhead do
15	13	2.4	540	17	W03762	Interleukin 13
14	15	2.2	106	20	Y55730	FKHL8 forkhead dom
15	14	2.2	106	20	Y55731	FKHL17 forkhead do
16	12	2.2	106	20	Y55732	FKHL13 forkhead do
17	12	2.2	106	20	Y55738	Nephila clavipes s
18	11	2.0	37	19	W53373	Nephila clavipes s
19	11	2.0	37	19	W53372	Nephila clavipes s
19	11	2.0	37	21	Y59098	N. clavipes spider
20	11	2.0	37	21	Y59099	N. clavipes spider
21	11	2.0	531	12	Y40098	Spider silk protel
22	11	2.0	595	12	R43099	N.clavipes draglin
23	11	2.0	595	19	W53347	Nephila clavipes s
24	11	2.0	595	21	Y59071	N. clavipes spider
25	11	2.0	921	21	Y77293	Streptomyces dradi
26	11	2.0	921	21	Y77293	KegAtq loading di
27	11	2.0	922	21	W28845	Hydrolase synthase
28	11	2.0	4472	18	W22601	Hybrid smg/LyG O
29	11	2.0	4455	18	W22611	Platelet glycoprob
30	10	1.8	14	20	W67765	Mouse beta-actin a
31	10	1.8	15	20	W92610	Platelet glycoprob
32	10	1.8	15	20	W67766	peptide derived fr
33	10	1.8	18	18	W32212	peptide linker uses
34	10	1.8	19	21	R83221	Platelet glycoprot
35	10	1.8	23	20	W67767	Platelet glycoprot
36	10	1.8	24	19	W79139	GG24 Gly-ala inser
37	10	1.8	35	16	R80211	Major ampullate sp
38	10	1.8	50	18	W13632	Apollipoprotein (a)
39	10	1.8	50	18	W13633	Apollipoprotein (a)
40	10	1.8	92	21	V44712	Shortorn sculpin
41	10	1.8	106	20	Y55727	FKHL15 forkhead do
42	10	1.8	106	20	Y55728	FKHL5 forkhead dom
43	10	1.8	106	20	Y55732	FKHL6 forkhead dom
44	10	1.8	106	20	Y55733	Spider dragline ve
45	10	1.8	119	16	R99058	

## ALIGNMENTS

PS Claim 3; Fig 2; 65pp; English.

CC The invention relates to the use of mutations in the FRCAC3 gene for  
CC detecting predisposition to diseases of the eye or developmental defects  
CC and for developing products for prevention and therapy. The method  
CC comprises analyzing nucleic acid of the mammal to determine whether the  
CC nucleic acid contains a mutation in a FRCAC3 gene, where the presence of  
CC the mutation is an indication that the mammal has an increased likelihood  
CC of developing a disease of the eye. The methods can be used for the  
CC diagnosis of increased likelihood of developing a disease of the eye,  
CC e.g., glaucoma, or a developmental defect. The methods can also be used  
CC for the prevention and treatment of diseases of the eye. Transgenic  
CC animals containing the mutated FRCAC3 gene can be used to screen for  
CC therapeutic compounds that modulate anterior segment dysgenesis or the  
CC onset or progression of glaucoma via a FRCAC3-dependent or FRCAC3-  
CC affected pathway. The present sequence represents the FRCAC3 polypeptide.

Sequence 553 AA;

Query Match 100.0%; Score 553; DB 20; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOARVSVSPNSLGVVPTLGEQSYRAAAAAGGCTAMAPMSVYSHPAHQEPGGM 60  
DB 1 mgarysvspnslgvvpyl9geqsyraaaaaaggytampamsvyshpahea9y9gm 60  
QY 61 ARAYGTYTPQOPKDMWKPYSYIALITWAIQNAPODKITLNGIYOFIMDRPFRONKQ 120  
DB 61 ataygtytpqopkdmwkpysyialitwainapdkkitlningyofimdrfpyrdnkq 120  
QY 121 GMONSTRHNLSTNEGVKVPKRDCKKPGKSTWTLDPDSTNMFENGSLRRRRRRKKKDAL 180  
DB 121 gmonstrhnlstnegvkvpkrdckkpgkswtldpdsymfengsflrrrrfkkkda 180  
QY 181 KKEKEDRLHKEPPPPASPRPAPPEADGNAPOPPPRRIODIKTENGTCSPPQPL 240  
DB 181 kkeekedrlhkeppppasprpapeadgnapppprvridikktengtcpsppql 240  
QY 241 SPAALGSGSAAVAKTISPPSSSSLSGSSPPGSLPSARPLSLDGADAPPPAPAP 300  
DB 241 spaalsgsgsaaavaktisppsssslssgssppgsllpsarplsldgadsappppasap 300  
QY 301 PPHHSQGFVNDIMTSLGSPQSAAEISGLLASAASSRAGIAPPLALGAYSPQSSSL 360  
DB 301 pphhsqgfsvndimtslrgspqsaaelsgllasaastraglapplalqayspqsssl 360  
QY 361 YSPSPQSOTSSAGSGGGGGAAGAGAGAGTYHCNLIQAMSLYAAGERGHLQAGAGAG 420  
DB 361 yspspqsotssagsgggggaagagagagtyhcnliqamslyaagerghlqagagag 420  
QY 421 GSAVDDPLPYSLPPVYSSSSSSLSHGGGGGGGGGGGDEGHHRAHQGRILSWYLQACG 480  
DB 421 gsavddplpyslppvyssssslsshggggggggggdeghhpaahgrilswylngaag 480  
QY 481 DLGHLASAAAATAAAGPGCOQNFHVSREMFESQRTGLNNSPVNGSSCOMAFPSOSLY 540  
DB 481 dlghlasaaaataaagpgcoqnfhvsremfesqrlglnnspvngsscomafpsosly 540  
QY 541 RTSGARFYDCKSKF 553  
DB 541 rtsgarfvydcskf 553

RESULT 2  
Y55721  
ID Y55721 standard; Protein: 553 AA.

AC Y55721;  
XX  
DT 22-FEB-2000 (first entry)  
XX

DE Human forkhead transcription factor gene, FKHL7 protein sequence.  
XX  
KW Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
KM transgenic animal; drug screening.  
XX  
OS Homo sapiens.  
XX

PN W0953060-A2.

XX 21-OCT-1999.

PD 14-APR-1999; 99WO-US08148.

PF 15-APR-1998; 98US-0081870.

PR 22-MAY-1998; 98US-0083352.

XX (TOWA ) UNIV IOWA RES FOUND.

PA Sheffield VC, Alward MLM, Stone EM, Nishimura D, Patel S;

XX MPI: 1999-620429/53.

DR N-PSDB; 238079, 238080.

XX

PS Claim 7; Fig 1; 99pp; English.

CC The invention provides a human forkhead transcription factor gene, FKHL7.  
CC The FKHL7 protein can be produced by standard recombinant methodology.  
CC The products can be used for diagnosis, prognosis, monitoring, prevention  
CC and treatment of glaucoma. They can also be used for the production of  
CC transgenic animals and drug screening. The present sequence represents  
CC the protein sequence of the FKHL7 gene.

Sequence 553 AA;

Query Match 45.2%; Score 250; DB 20; Length 553;  
Best Local Similarity 99.7%; Pred. No. 4.3e-209;  
Matches 350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 203 PAPPQADGNAPGQPPPPVRIODIKTENGTCPSPQPLSPFAALGSGSAAVAKTISPD 262  
DB 203 pappqadgnapgpqpppvritdikktengtcpsppqplspfaalgsaaavaktiespd 262  
QY 263 SSSLSGSSSPGSLPSARPLSLDGADAPPPAPSPHHSQGFVNDIMTSLRGSPQ 322  
DB 263 ssslssgssspgslpsarplsldgadsappppaspphhsqgfsvndimtslrgspq 322  
QY 323 SAAELSSGLASAAASRRAGIAPPLALGAYSPGOSLYSSPCSTSSAGSGGGGGGAG 382  
DB 323 saaelslsgllsaaasrraglapplalqayspgqsllyspcsqtsaagsggggag 382  
QY 383 AAGAGAGAGTYHCNLIQAMSLYAAGERGHLQAGAGGSAVDDPLPYSLPPVYSSSS 442  
DB 383 aagagagagtyhcnliqamslyaagerghlqagaggsavddplpyslppvysss 442  
QY 443 SLSHGGGGGGGGGDEGHHRAHQGRILSWYLQAGDGLHLASAAAATAAAGPGCOQ 502  
DB 443 slshggggggggdeghhpaahgrilswylngaagdglhlasaaaataaagpgcoq 502  
QY 503 NFHSVREMFESQRTGLNNSPVNGSSCOMAFPSOSLYRTSGARFYDCKSKF 553  
DB 503 nfhsvremfesqrlglnnspvngsscomafpsoslyrtsgarfvydcskf 553

RESULT 3  
Y43260  
ID Y43260 standard; Protein: 553 AA.  
AC Y43260;  
XX  
DT 22-FEB-2000 (first entry)  
XX



DT 17-JAN-2000 (first entry)  
 DE Human FKHL7 protein sequence.  
 XX  
 XX FKHL7, human; forkhead transcription factor gene; diagnosis; therapy;  
 KW congenital heart disease.  
 XX  
 XX Homo sapiens.  
 OS  
 XX MO952415-A2.  
 PN  
 PD 21-OCT-1999.  
 XX  
 XX 14-APR-1999; 99MO-US08159.  
 PF  
 XX 15-APR-1998; 98US-0081870.  
 PR 22-MAY-1998; 98US-0083351.  
 XX  
 XX (IOWA ) UNIV IOWA RES. FOUNDED.  
 PA  
 XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patel S;  
 PI  
 XX WPI: 1999-620257/53.  
 DR N-PSDB: 231671, 231672.  
 XX  
 XX New isolated human forkhead transcription factor gene, FKHL7, used to  
 PT develop products for the diagnosis, prognosis, monitoring, prevention  
 PT or treatment of congenital heart disease  
 PS  
 XX Claim 17; Fig 1; 98pp; English.  
 CC This sequence represents the human forkhead transcription factor gene,  
 CC designated FKHL7, of the invention. FKHL7 can be used in a novel method  
 CC for treating or preventing the development of a congenital heart disease  
 CC (CHD) in a subject. The FKHL7 sequences can be used for diagnosis,  
 CC prognosis, monitoring, prevention and treatment of CHD. They can also be  
 CC used for the production of transgenic animals and drug screening.  
 CC  
 XX Sequence 553 AA;  
 SQ  
 Query Match 40.3%; Score 223; DB 20; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-185; Indels 0; Gaps 0;  
 Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 203 PAPERQADGNARPGQPPRVRIODIKTEKNGTCSPPQPLSPAAALGSGSAAPVKTIESPDS 262  
 DB 203 papeqadgnarpqprrvridikentgcspqpplspaaalgsgsaavpkiespds 262  
 QY 263 SSSSLSSGSSPPGSLPSPARPLSDGADSPAPRPAPRPHHSGQSVNIMTSLRGSPQ 322  
 DB 263 sssslssgssppgslpsparplsdgadspaprpaprhhsqgsvnintslrgspq 322  
 QY 323 SAAALSSGLASAASRAGIAPPLALGAYSPGSSLYSSPQSSQTSAGSGSGGCGGCG 382  
 DB 323 saaalssgllasaasragiapplalgayspgsslysspsqtsagsgsgcgcgcg 382  
 QY 383 AAGGAGAGACTYHCNLQAMSLYAGERGHLOGAGGAGSAYD 425  
 DB 383 aaggagagactyhcnlqamslyaagerghlogaggagsayd 425  
 RESULT 4  
 Y01097  
 ID Y01097 standard; Protein: 501 AA.  
 XX  
 XX Y01097;  
 AC  
 XX 11-JUN-1999 (first entry)  
 DT  
 XX FKHL7 protein.  
 DE  
 XX FKHL7, human; forkhead transcription factor gene; diagnosis; therapy;  
 KW congenital heart disease.  
 XX  
 XX Homo sapiens.  
 OS  
 XX MO952415-A2.  
 PN  
 PD 21-OCT-1999.  
 XX  
 XX 14-APR-1999; 99MO-US08159.  
 PF  
 XX 15-APR-1998; 98US-0081870.  
 PR 22-MAY-1998; 98US-0083351.  
 XX  
 XX (IOWA ) UNIV IOWA RES. FOUNDED.  
 PA  
 XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patel S;  
 PI  
 XX WPI: 1999-620257/53.  
 DR N-PSDB: 231671, 231672.  
 XX  
 XX New isolated human forkhead transcription factor gene, FKHL7, used to  
 PT develop products for the diagnosis, prognosis, monitoring, prevention  
 PT or treatment of congenital heart disease  
 PS  
 XX Claim 17; Fig 1; 98pp; English.  
 CC This sequence represents the human forkhead transcription factor gene,  
 CC designated FKHL7, of the invention. FKHL7 can be used in a novel method  
 CC for treating or preventing the development of a congenital heart disease  
 CC (CHD) in a subject. The FKHL7 sequences can be used for diagnosis,  
 CC prognosis, monitoring, prevention and treatment of CHD. They can also be  
 CC used for the production of transgenic animals and drug screening.  
 CC  
 XX Sequence 553 AA;  
 SQ

KW lipid metabolism; adipocyte differentiation; obesity-related condition;  
 KW non-insulin dependent diabetes mellitus; cardiovascular disease;  
 KW catabolic condition; anorexia; bulimia; therapy.  
 XX  
 XX Homo sapiens.  
 OS  
 XX MO9854216-A1.  
 PN  
 PD 03-DEC-1998.  
 XX  
 XX 26-MAY-1998; 98MO-SF00989.  
 PF  
 XX 26-MAY-1997; 97SF-0001963.  
 PR  
 XX (PHAA ) PHARMACIA & UPJOHN AB.  
 PA  
 XX Carlsson P, Enerbaeck S;  
 PI  
 XX WPI: 1999-070211/06.  
 DR N-PSDB: X28103.  
 XX  
 XX New transcription factor designated FREAC11 - which regulates  
 PT adipose tissue expressed genes involved in lipid metabolism and  
 PT adipocyte differentiation, used to, e.g. inhibit gene expression  
 PS  
 XX Claim 5; Fig 1; 23pp; English.  
 CC This sequence represents the FREAC11 protein of the invention. The  
 CC protein has transcriptional regulatory function directed against adipose  
 CC tissue expressed genes. The protein is involved in lipid metabolism  
 CC and/or adipocyte differentiation. FREAC11 is selectively expressed in  
 CC adipose tissue, and is used for transcriptional regulation of adipocyte  
 CC expressed genes. The DNA or protein can be used as a drug that blocks  
 CC obesity-related conditions to increase or decrease the activity of  
 CC the cis elements of adipose expressed genes, e.g. FREAC11 can be used as a drug that blocks  
 CC the cis elements of adipose expressed genes to inhibit gene expression.  
 CC The FREAC11 protein can also be used for high throughput screening for  
 CC substances that affect the activity of FREAC11, such as inhibitors,  
 CC antagonists, or agonists. Conditions which can be treated include  
 CC obesity, non-insulin dependent diabetes mellitus, cardiovascular  
 CC diseases, catabolic conditions, anorexia, bulimia.  
 CC  
 XX Sequence 501 AA;  
 SQ  
 Query Match 11.0%; Score 61; DB 20; Length 501;  
 Best Local Similarity 100.0%; Pred. No. 6e-45; Indels 0; Gaps 0;  
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 118 NKGQMNSTRHNLINIECFVAVPRDQKRGKSTWTLDPDSYNNFENGSLRRRRFRKK 177  
 DB 112 nkqgwnstrhnlslneicfvavprdkrgkstwtldpdsynnfnegslrrrrfrkk 171  
 QY 178 D 178  
 DB 172 d 172  
 RESULT 5  
 Y55723  
 ID Y55723 standard; Protein: 106 AA.  
 XX  
 XX Y55723;  
 AC  
 XX 22-FEB-2000 (first entry)  
 DT  
 XX FKHL7 forkhead domain fragment.  
 DE  
 XX Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
 KW transgenic animal; drug screening.  
 XX  
 XX Homo sapiens.  
 OS

XX PN WO9953060-A2.  
 XX PD 21-OCT-1999.  
 XX PF 14-APR-1999; 99WO-US08148.  
 XX PR 15-APR-1998; 98US-0081870.  
 XX PR 22-MAY-1998; 98US-0083352.  
 XX PA (IOWA ) UNIV IOWA RES FOUND.  
 XX PI Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 XX DR WPI; 1999-620429/53.  
 XX PT New isolated human forkhead transcription factor gene, FKHL7, used to,  
 XX PT e.g. develop products for the diagnosis -  
 XX PS Disclosure; Fig 2; 99pp; English.  
 CC The invention provides a human forkhead transcription factor gene, FKHL7.  
 CC The FKHL7 protein can be produced by standard recombinant methodology.  
 CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 CC and treatment of glaucoma. They can also be used for the production of  
 CC transgenic animals and drug screening. Sequences Y55722-739 represent  
 CC forkhead domains of different members of the FKHL-family of genes.  
 CC XX  
 SQ Sequence 106 AA;

Query Match  
 Best Local Similarity 8.7%; Score 48; DB 20; Length 106;  
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 90 AIONAPDKKTLNCTYGFIDRPEFFYRDNKGQNSIRHNLINCEFY 137  
 Db 18 aignapdkkittlingyqfndripdyrdkqgwsairhnlslnectf 65

RESULT 6

ID Y55724 standard; Protein; 106 AA.  
 AC Y55724;  
 DT 22-FEB-2000 (first entry)  
 DE FKHL14 forkhead domain fragment..  
 XX KW Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
 XX KW transgenic animal; drug screening.  
 XX OS Unidentified.  
 XX PN WO9953060-A2.  
 XX PD 21-OCT-1999.  
 XX PF 14-APR-1999; 99WO-US08148.  
 XX PR 15-APR-1998; 98US-0081870.  
 XX PR 22-MAY-1998; 98US-0083352.  
 XX PA (IOWA ) UNIV IOWA RES FOUND.  
 XX PI Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 XX DR WPI; 1999-620429/53.  
 XX PT New isolated human forkhead transcription factor gene, FKHL7, used to,  
 XX PT e.g. develop products for the diagnosis -

PS Disclosure; Fig 2; 99pp; English.  
 XX CC The invention provides a human forkhead transcription factor gene, FKHL7.  
 XX CC The FKHL7 protein can be produced by standard recombinant methodology.  
 XX CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 XX CC and treatment of glaucoma. They can also be used for the production of  
 XX CC transgenic animals and drug screening. Sequences Y55722-739 represent  
 XX CC forkhead domains of different members of the FKHL-family of genes.  
 XX XX  
 SQ Sequence 106 AA;

Query Match  
 Best Local Similarity 7.2%; Score 40; DB 20; Length 106;  
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 139 VPRDDKKPKGKSYWTLDPDPSYMFENGSELRRRRRFKKDD 178  
 Db 67 vprddkkpkygswtldpdsymfengsflrrrrfkdkd 106

RESULT 7

ID Y55725 standard; Protein; 106 AA.  
 AC Y55725;  
 DT 22-FEB-2000 (first entry)  
 DE FKHL18 forkhead domain fragment.  
 XX KW Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
 XX KW transgenic animal; drug screening.  
 XX OS Unidentified.  
 XX PN WO9953060-A2.  
 XX PD 21-OCT-1999.  
 XX PF 14-APR-1999; 99WO-US08148.  
 XX PR 15-APR-1998; 98US-0081870.  
 XX PR 22-MAY-1998; 98US-0083352.  
 XX PA (IOWA ) UNIV IOWA RES FOUND.  
 XX PI Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 XX DR WPI; 1999-620429/53.  
 XX PT New isolated human forkhead transcription factor gene, FKHL7, used to,  
 XX PT e.g. develop products for the diagnosis -  
 XX PS Disclosure; Fig 2; 99pp; English.  
 CC The invention provides a human forkhead transcription factor gene, FKHL7.  
 CC The FKHL7 protein can be produced by standard recombinant methodology.  
 CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 CC and treatment of glaucoma. They can also be used for the production of  
 CC transgenic animals and drug screening. Sequences Y55722-739 represent  
 CC forkhead domains of different members of the FKHL-family of genes.  
 CC XX  
 SQ Sequence 106 AA;

Query Match  
 Best Local Similarity 3.1%; Score 17; DB 20; Length 106;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 121 GWQNSIRHNLINCEFY 137  
 Db 49 gwqnsirhnlslnectf 65

RESULT 8  
ID Y55735 standard; Protein; 105 AA.  
XX Y55735;  
AC 22-FEB-2000 (first entry)  
XX  
XX  
DE FKHL2 forkhead domain fragment.  
XX Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
KM transgenic animal; drug screening.  
XX  
XX Unidentified.  
XX  
XX WO953060-A2.  
XX  
XX 21-OCT-1999.  
XX  
XX 14-APR-1999; 99WO-US08148.  
XX  
XX 15-APR-1998; 98US-0081870.  
XX 22-MAY-1998; 98US-0083352.  
XX  
XX (IOWA ) UNIV IOWA RES FOUND.  
XX  
XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
PI WPI: 1999-620429/53.  
XX  
XX WPI: 1999-620429/53.  
XX  
XX New isolated human forkhead transcription factor gene, FKHL7, used to,  
PT e.g. develop products for the diagnosis -  
XX  
XX Disclosure; Fig 2; 99pp; English.  
XX  
XX The invention provides a human forkhead transcription factor gene, FKHL7.  
XX The FKHL7 protein can be produced by standard recombinant methodology.  
XX The products can be used for diagnosis, prognosis, monitoring, prevention  
XX and treatment of glaucoma. They can also be used for the production of  
XX transgenic animals and drug screening. Sequences Y55722-739 represent  
XX forkhead domains of different members of the FKHL-family of genes.  
XX  
XX Sequence 105 AA:  
SQ

Query Match 2.9%; Score 16; DB 20; Length 105;  
Best Local Similarity 100.0%; Pred. No. 1.9e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKQSGWNSIRHNLN 133  
DB 46 nkqsgwnsirhnlsln 61

RESULT 9  
ID Y55726 standard; Protein; 106 AA.  
XX Y55726;  
AC 22-FEB-2000 (first entry)  
XX  
XX  
DE FKHL1 forkhead domain fragment.  
XX Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
KM transgenic animal; drug screening.  
XX  
XX Unidentified.  
XX  
XX WO953060-A2.  
XX

PD 21-OCT-1999.  
XX  
XX 14-APR-1999; 99WO-US08148.  
XX  
XX 15-APR-1998; 98US-0081870.  
XX 22-MAY-1998; 98US-0083352.  
XX  
XX (IOWA ) UNIV IOWA RES FOUND.  
XX  
XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
PI WPI: 1999-620429/53.  
XX  
XX WPI: 1999-620429/53.  
XX  
XX New isolated human forkhead transcription factor gene, FKHL7, used to,  
PT e.g. develop products for the diagnosis -  
XX  
XX Disclosure; Fig 2; 99pp; English.  
XX  
XX The invention provides a human forkhead transcription factor gene, FKHL7.  
XX The FKHL7 protein can be produced by standard recombinant methodology.  
XX The products can be used for diagnosis, prognosis, monitoring, prevention  
XX and treatment of glaucoma. They can also be used for the production of  
XX transgenic animals and drug screening. Sequences Y55722-739 represent  
XX forkhead domains of different members of the FKHL-family of genes.  
XX  
XX Sequence 106 AA:  
SQ

Query Match 2.9%; Score 16; DB 20; Length 106;  
Best Local Similarity 100.0%; Pred. No. 1.9e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 TLNGYGFIMDRPPY 115  
DB 28 tlngygfimdrpfy 43

RESULT 10  
ID Y55734 standard; Protein; 106 AA.  
XX Y55734;  
AC 22-FEB-2000 (first entry)  
XX  
XX  
DE FKHL4 forkhead domain fragment.  
XX Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
KM transgenic animal; drug screening.  
XX  
XX Unidentified.  
XX  
XX WO953060-A2.  
XX  
XX 21-OCT-1999.  
XX  
XX 14-APR-1999; 99WO-US08148.  
XX  
XX 15-APR-1998; 98US-0081870.  
XX 22-MAY-1998; 98US-0083352.  
XX  
XX (IOWA ) UNIV IOWA RES FOUND.  
XX  
XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
PI WPI: 1999-620429/53.  
XX  
XX WPI: 1999-620429/53.  
XX  
XX New isolated human forkhead transcription factor gene, FKHL7, used to,  
PT e.g. develop products for the diagnosis -  
XX  
XX Disclosure; Fig 2; 99pp; English.  
XX  
XX The invention provides a human forkhead transcription factor gene, FKHL7.

CC The FKHL7 protein can be produced by standard recombinant methodology.  
 CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 CC and treatment of glaucoma. They can also be used for the production of  
 CC transgenic animals and drug screening. Sequences Y55722-739 represent  
 CC forkhead domains of different members of the FKHL-family of genes.  
 XX  
 SQ Sequence 106 AA;

Query Match 2.9%; Score 16; DB 20; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKOGMNSIRHNSLN 133  
 Db 46 nkqgwnsirhnlsln 61

RESULT 11  
 R44551  
 ID R44551 standard; Protein; 480 AA.  
 XX  
 AC R44551;

DT 02-JUN-1994 (first entry)  
 XX  
 DE Brain factor-1.

XX Brain factor; BF-1; BF-2; BF-3; DNA binding domain;  
 KW brain transcription factor; diagnosis; tumour; cancer; probe;  
 telencephalon.  
 XX  
 OS Rattus rattus.

XX  
 FH Key Location/Qualifiers  
 FT Domain 162..271  
 FT /Label= DNA\_binding\_domain\_homology  
 XX  
 FN W09323430-A.

XX 25-NOV-1993.  
 PD  
 XX

PE 30-APR-1993; 93MO-US04102.  
 XX

PR 13-MAY-1992; 92US-0882292.  
 XX

PA (SLOK ) SLOAN KETTERING INST CANCER.  
 XX

PL Lai E, Tao W;  
 XX

DR WPI: 1993-386481/48.  
 DR N-PSDB; Q50631.  
 XX

PT Isolated, animal nucleic acids encoding brain transcription  
 PT factors - useful for diagnosis and treatment of abnormal brain  
 PT factor synthesis in tumour tissue from animals and utilised as  
 PT probe  
 XX  
 PS Claim 23; Page 57-60; 96pp; English.

CC The brain factors are transcription factors. Abnormal expression of  
 CC BF-1 in telencephalon-derived tissue or tumour tissue can be  
 CC diagnosed. BF DNA or protein can be used to correct defective  
 CC synthesis of BF.  
 CC  
 XX

SO Sequence 480 AA;

Query Match 2.9%; Score 16; DB 14; Length 480;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKOGMNSIRHNSLN 133

Db 212 nkqgwnsirhnlsln 227

RESULT 12  
 Y55737  
 ID Y55737 standard; Protein; 76 AA.  
 XX  
 AC Y55737;

DT 22-FEB-2000 (first entry)  
 XX  
 DE FKHL10 forkhead domain partial fragment.

XX Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
 KW transgenic animal; drug screening.  
 XX  
 OS Unidentified.

PN W09953060-A2.  
 XX

PD 21-OCT-1999.  
 XX

PE 14-APR-1999; 99MO-US08148.  
 XX

PR 15-APR-1998; 98US-0081870.  
 PR 22-MAY-1998; 98US-0083352.  
 XX

PA (IOWA ) UNIV IOWA RES FOUND.  
 XX

PI Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 XX

DR WPI: 1999-620429/53.  
 XX

PT New isolated human forkhead transcription factor gene, FKHL7, used to,  
 PT e.g. develop products for the diagnosis -  
 XX

PS Disclosure; Fig 2; 99pp; English.  
 XX

CC The invention provides a human forkhead transcription factor gene, FKHL7.  
 CC The FKHL7 protein can be produced by standard recombinant methodology.  
 CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 CC and treatment of glaucoma. They can also be used for the production of  
 CC transgenic animals and drug screening. Sequences Y55722-739 represent  
 CC forkhead domains of different members of the FKHL-family of genes.  
 XX

SO Sequence 76 AA;

Query Match 2.4%; Score 13; DB 20; Length 76;  
 Best Local Similarity 100.0%; Pred. No. 0.00057;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GWMNSIRHNSLN 133  
 Db 49 gwmsirhnlsln 61

RESULT 13  
 Y55729  
 ID Y55729 standard; Protein; 106 AA.  
 XX  
 AC Y55729;

DT 22-FEB-2000 (first entry)  
 XX  
 DE FKHL9 forkhead domain fragment.

XX Forkhead transcription factor gene; FKHL7; treatment; glaucoma; human;  
 KW transgenic animal; drug screening.  
 XX  
 OS Unidentified.

XX

PN W09953060-A2.  
 XX 21-OCT-1999.  
 XX 14-APR-1999; 99WO-US08148.  
 XX 15-APR-1998; 98US-0081870.  
 PR 22-MAY-1998; 98US-0083352.  
 XX (IOWA ) UNIV IOWA RES FOUND.  
 XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 PI WPI: 1999-620429/53.  
 DR  
 XX  
 PT New isolated human forkhead transcription factor gene, FKHL7, used to,  
 PT e.g. develop products for the diagnosis -  
 XX  
 PS Disclosure: Fig 2; 99pp; English.  
 CC The invention provides a human forkhead transcription factor gene, FKHL7.  
 CC The FKHL7 protein can be produced by standard recombinant methodology.  
 CC The products can be used for diagnosis, prognosis, monitoring, prevention  
 CC and treatment of glioma. They can also be used for the production of  
 CC transgenic animals and drug screening. Sequences Y5722-739 represent  
 CC forkhead domains of different members of the FKHL-family of genes.  
 CC  
 XX Sequence 106 AA;  
 SQ

Query Match 2.4%; Score 13; DB 20; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 0.00076;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 121 GWNSTRHNLIN 133  
 |||||  
 Db 49 gwnstrhnlins 61

RESULT 14  
 W03762  
 ID W03762 standard; Protein: 540 AA.  
 XX  
 AC W03762;  
 XX  
 DT 31-OCT-1996 (first entry)  
 XX  
 DE Interleukin binding factor ILF-1.  
 XX  
 KW Interleukin binding factor; ILF-1; purine rich motif; HIV-LTR;  
 KW human immunodeficiency virus; long terminal repeat; fork head;  
 KW AIDS; DNA binding; nuclear factor of activated T cell; NFAT;  
 KW autoimmune disease; rheumatism.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT Binding-site  
 FT Location/Qualifiers  
 FT 96..101  
 FT /label= nucleotide-binding\_site  
 FT /note= "putative; is homologous to sequences  
 FT found in a number of enzymes including  
 FT the ras oncogene"  
 FT 228..233  
 FT /label= nuclear\_localisation\_signal  
 FT /note= "putative"  
 FT 239..247  
 FT /label= ubiquitin-mediated\_degradation\_signal  
 FT /note= "putative"  
 FT 458..460  
 FT /label= N-glycosylation\_site  
 FT /note= "putative"  
 FT 136..233  
 FT /note= "homologous to Drosophila fork head DNA-  
 FT Domain

binding domain"  
 XX US5534631-A.  
 XX 09-JUL-1996.  
 XX 30-JUN-1992; 92US-0906930.  
 XX 30-JUN-1992; 92US-0906930.  
 XX (TEXA ) UNIV TEXAS SYSTEM.  
 XX Gaynor RB, Li C, Nirula A;  
 PI WPI: 1996-333310/33.  
 DR N-PSDB: T36079.  
 DR  
 XX  
 PT New DNA encoding interleukin (IL) nucleic acid binding protein -  
 PT that binds to purine-rich motifs in viral and IL gene promoters to  
 PT modulate their expression, e.g. for treating AIDS and auto-immune  
 PT disease  
 PS Claim 3: Columns 29-32; 61pp; English.  
 XX  
 CC The present sequence is that of interleukin binding factor ILF-1,  
 CC which binds specifically to a purine-rich motif (the NFAT binding  
 CC site) in the long terminal repeat of HIV and, with lower affinity,  
 CC to a similar sequence in the interleukin (IL)-2 promoter. Database  
 CC searches have shown that a 98 amino acid region of ILF-1 is  
 CC homologous to the Drosophila fork head regulatory protein, and the  
 CC DNA binding domain of the hepatocyte specific factor HNF-3A. The  
 CC ILF regulates IL-2 expression in T lymphocytes and is  
 CC involved in positive and negative regulation of viral and cellular  
 CC genes; it is thus potentially useful for treating AIDS and to  
 CC inhibit lymphocytes in rheumatic and other autoimmune diseases.  
 CC A second ILF protein, designated ILF-2, has also been identified.  
 CC The two proteins are identical up to amino acid 475 and then differ  
 CC in their C-terminal sequences.  
 CC  
 XX Sequence 540 AA;  
 SQ

Query Match 2.4%; Score 13; DB 17; Length 540;  
 Best Local Similarity 100.0%; Pred. No. 0.0032;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 121 GWNSTRHNLIN 133  
 |||||  
 Db 181 gwnstrhnlins 193

RESULT 15  
 Y5730  
 ID Y5730 standard; Protein: 106 AA.  
 XX  
 AC Y5730;  
 XX  
 DT 22-FEB-2000 (first entry)  
 XX  
 DE FKHL8 forkhead domain fragment.  
 XX  
 KW Forkhead transcription factor gene; FKHL7; treatment; glioma; human;  
 KW transgenic animal; drug screening.  
 XX  
 OS Unidentified.  
 XX  
 PN W09953060-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 14-APR-1999; 99WO-US08148.  
 XX  
 PR 15-APR-1998; 98US-0081870.

PR 22-MAY-1998; 98US-0083352.  
XX  
XX  
PA (IOWA ) UNIV IOWA RES FOUNO.  
XX  
XX  
PI Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
XX  
XX WPI, 1999-620429/53.  
DR  
XX  
XX  
PT New isolated human forkhead transcription factor gene, FKHL7, used to,  
PT e.g. develop products for the diagnosis -  
XX  
XX  
PS Disclosure: Fig 2; 99pp; English.  
XX  
XX  
CC The invention provides a human forkhead transcription factor gene, FKHL7.  
CC The FKHL7 protein can be produced by standard recombinant methodology.  
CC The products can be used for diagnosis, prognosis, monitoring, prevention  
CC and treatment of glaucoma. They can also be used for the production of  
CC transgenic animals and drug screening. Sequences Y55722-739 represent  
CC forkhead domains of different members of the FKHL-family of genes.  
XX  
XX  
SQ Sequence 106 AA;

Query Match 2.2%; Score 12; DB 20; Length 106;  
Best Local Similarity 100.0%; Pred. No. 0.0056;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 77 VKPPSYIALIT 88  
|||  
Db 5 vkppsyialit 16

Search completed: February 6, 2001, 22:52:58  
Job time: 8827 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 6, 2001, 21:50:24 ; Search time 60.6 Seconds  
(without alignments) 1069.571 Million cell updates/sec

Title: US-09-292-862-2  
Perfect score: 553  
Sequence: 1 MQARRSVSSPNSLGVPVILG.....PSSQSLVRSQAFVYDCSKF 553

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 374700 seqs, 117207915 residues

Word size: 0

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

SPREMBL\_15:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.rodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	351	63.5	553	4 Q9NUE5	Q9NUE5 homo sapien
2	120	21.7	553	11 Q9QWR9	Q9QWR9 mus musculu
3	69	12.5	528	13 Q93440	Q93440 gallus gall
4	68	12.3	492	13 Q9YHB2	Q9YHB2 xenopus lae
5	68	12.3	492	13 Q9PVZ3	Q9PVZ3 xenopus lae
6	61	11.0	445	13 Q9Y771	Q9Y771 gallus gall
7	61	11.0	445	13 Q13017	Q13017 gallus gall
8	40	7.2	461	13 Q9PVY8	Q9PVY8 xenopus lae
9	40	7.2	461	13 Q9PVY9	Q9PVY9 xenopus lae
10	39	7.1	108	4 Q9QDD0	Q9QDD0 homo sapien
11	21	3.8	270	5 Q22510	Q22510 caenorhabd
12	19	3.4	451	5 Q9VQV5	Q9VQV5 drosophila
13	16	2.9	62	13 Q9PS85	Q9PS85 xenopus xf
14	16	2.9	90	11 Q88470	Q88470 mus musculu
15	16	2.9	365	5 Q9VZK0	Q9VZK0 drosophila
16	16	2.9	402	5 Q61733	Q61733 brachydanio
17	16	2.9	420	13 Q73862	Q73862 brachydanio
18	16	2.9	436	13 Q9YHCS	Q9YHCS xenopus lae
19	15	2.7	202	5 P91278	P91278 caenorhabd

20	15	2.7	321	13 Q73784	Q73784 brachydanio
21	15	2.7	323	5 Q10924	Q10924 caenorhabd
22	15	2.7	352	13 Q9PRJ8	Q9PRJ8 xenopus lae
23	15	2.7	353	13 Q73785	Q73785 brachydanio
24	15	2.7	353	13 Q9PR68	Q9PR68 xenopus lae
25	15	2.7	371	13 Q73782	Q73782 brachydanio
26	15	2.7	443	13 P79770	P79770 gallus gall
27	15	2.7	456	5 Q9W1Y5	Q9W1Y5 drosophila
28	15	2.7	492	11 Q35392	Q35392 mus musculu
29	14	2.5	408	4 Q9NU39	Q9NU39 homo sapien
30	14	2.4	284	5 P91587	P91587 caenorhabd
31	13	2.4	322	5 Q9VQV4	Q9VQV4 drosophila
32	13	2.4	345	13 Q93613	Q93613 xenopus lae
33	13	2.4	345	13 Q9PSY4	Q9PSY4 xenopus lae
34	13	2.4	367	13 Q91905	Q91905 xenopus lae
35	13	2.4	370	13 Q91904	Q91904 xenopus lae
36	13	2.4	409	11 Q35939	Q35939 mus musculu
37	13	2.4	442	11 Q93133	Q93133 brachydanio
38	13	2.4	443	4 Q9UJL0	Q9UJL0 homo sapien
39	13	2.4	447	13 Q73628	Q73628 anolis caro
40	13	2.4	517	3 Q13606	Q13606 schizosach
41	13	2.4	586	3 P79006	P79006 schizosach
42	13	2.4	605	5 Q9YTF7	Q9YTF7 drosophila
43	13	2.4	609	4 Q13623	Q13623 homo sapien
44	13	2.4	642	3 Q60129	Q60129 schizosach
45	13	2.4	663	3 Q43058	Q43058 schizosach

## ALIGNMENTS

RESULT 1  
ID Q9NUE5 PRELIMINARY: PRT: 553 AA.  
AC Q9NUE5  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE D118B18.1 (FORKEHEAD BOX C1).  
GN FOXL1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Patel R.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AI034344; CAB6158.1; -  
DR SQUENCE 553 AA; 56788 MW; 5966P94303ED59A CRC64;  
SQ

Query Match	63.5%;	Score 351;	DB 4;	Length 553;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 351;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	203	PAPPEQADGNAPGPPPPVRIODIKTEGTCSPDPLSPAALGSGSAAVPIKIESPDS	262	
DB	203	PAPPEQADGNAPGPPPPVRIODIKTEGTCSPDPLSPAALGSGSAAVPIKIESPDS	262	
QY	263	SSSSISGSSPPGSLPSARPLSLDGADSDAPPPPAAPPPHSSOGFSVDNMTSLRGSPQ	322	
DB	263	SSSSISGSSPPGSLPSARPLSLDGADSDAPPPPAAPPPHSSOGFSVDNMTSLRGSPQ	322	
QY	323	SAAEELSSGLLAASAAASRRAGIAPPLALGAYSPGSSIVSSPQSTSSAGSSGGGGGAG	382	
DB	323	SAAEELSSGLLAASAAASRRAGIAPPLALGAYSPGSSIVSSPQSTSSAGSSGGGGGAG	382	
QY	383	AAGCAGGAGTYHCNLOAMSLYAGERGSHLQAGPAGGASAVDDPLPDVSLPPVSSSS	442	
DB	383	AAGCAGGAGTYHCNLOAMSLYAGERGSHLQAGPAGGASAVDDPLPDVSLPPVSSSS	442	
QY	443	SLSHGGGGGGGGGGGAGHHPAAGGRTLSWYLNQAGDGLHLASAAAAAGYPGQOQ	502	



DB 443 SLHHGGGGGGGGGGGAGHHPAAGRLTSLWYNAGGDLHLASAAAAAAGYPCOOQ 502  
 OY 503 NHSHVREMEESORIGLNNSPVNGNCCOMAPSSQSILYRTSGAFVYDCSKF 553  
 DB 503 NHSHVREMEESORIGLNNSPVNGNCCOMAPSSQSILYRTSGAFVYDCSKF 553

## RESULT 2

O9QWR9

ID O9QWR9 PRELIMINARY: PRT: 553 AA.

AC O9QWR9:

DT 01-MAY-2000 (TREMblrel. 13, Created)

DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)

GN FKHL/ME1

OS Mus musculus (Mouse)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=EMBRYO;

RA Hiemisch H., Schütz G., Kaestner K.H.;

RT "The mouse ~~chick~~ gene: cDNA sequence, chromosomal localization and

RL expression in adult tissues.";

EMBL: AJ23298; CA11239.1; -

HSSP: Q63245; 2HFH.

DR INTERPRO: IPR001766;

DR PFAM: PF00250; Fork head. 1.

DR PRINTS: PR00053; FORKHEAD.

DR PROSITE: PS00657; FORK\_HEAD\_1; 1.

DR PROSITE: PS00658; FORK\_HEAD\_2; 1.

DR PROSITE: PS00039; FORK\_HEAD\_3; 1.

SQ SEQUENCE 553 AA: 56939 MW: 3CDB14F69AA2F217 CRC64;

## Query Match

Best Local Similarity 100.0%; Score 120; DB 11; Length 553;

Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 60 MARAGPTPOPQPRDKWKKPYSYIALITMAIQAPPKKITLNGIYOFIMDRPFYRDK 119

DB 60 MARAGPTPOPQPRDKWKKPYSYIALITMAIQAPPKKITLNGIYOFIMDRPFYRDK 119

OY 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

DB 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

OY 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

DB 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

OY 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

DB 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

OY 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

DB 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

OY 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

DB 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

OY 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

DB 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

OY 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

DB 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

OY 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

DB 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

OY 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

DB 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

OY 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

DB 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

OY 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

DB 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

OY 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

DB 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

OY 120 OGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRRRFFKKDA 179

RT appendicular skeletal structures during chicken embryogenesis.";  
 RL Dev. Dyn. 212:94-101(1998).  
 DR EMBL: Y17794; CAAT6851.1; -  
 DR HSSP: Q63245; 2HFH.  
 DR INTERPRO: IPR001766; -  
 DR PFAM: PF00250; Fork head. 1.  
 DR PRINTS: PR00053; FORKHEAD.

## RESULT 4

O9YHB2

ID O9YHB2 PRELIMINARY: PRT: 492 AA.

AC O9YHB2:

DT 01-MAY-1999 (TREMblrel. 10, Created)

DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)

DE 01-MAY-2000 (TREMblrel. 13, Last annotation update)

GN WINGED HELIX TRANSCRIPTION FACTOR XFD-11

OS Xenopus laevis (African clawed frog)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

OX NCBI\_TaxID=8135;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=98440281; PubMed=9767159;

RA Koster M., Dillinger K., Knoche W.;

RT "Expression pattern of the winged helix factor XFD-11 during Xenopus

RL embryogenesis.";

EMBL: AF116844; AAC9469.1; -

HSSP: Q63245; 2HFH.

DR INTERPRO: IPR001766; -

DR PFAM: PF00250; Fork head. 1.

DR PRINTS: PR00053; FORKHEAD.

DR PROSITE: PS00657; FORK\_HEAD\_1; 1.

DR PROSITE: PS00658; FORK\_HEAD\_2; 1.

DR PROSITE: PS00039; FORK\_HEAD\_3; 1.

SQ SEQUENCE 492 AA: 53643 MW: A6C66458CF44465F CRC64;

## Query Match

Best Local Similarity 100.0%; Score 68; DB 13; Length 492;

Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 RPFYRDNKOGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRR 170

DB 112 RPFYRDNKOGWONSIRHNLSTNECFVKKVPRDCKKPGKGSYWTLPDPSYNNFENGSLRR 171

OY 171 RRRRKKD 178

DB 172 RRRRKKD 179

OY 171 RRRRKKD 178

DB 172 RRRRKKD 179

OY 171 RRRRKKD 178

DB 172 RRRRKKD 179

OY 171 RRRRKKD 178

DB 172 RRRRKKD 179

OY 171 RRRRKKD 178

DB 172 RRRRKKD 179

OY 171 RRRRKKD 178

DB 172 RRRRKKD 179

OY 171 RRRRKKD 178

DB 172 RRRRKKD 179

OY 171 RRRRKKD 178

DB 172 RRRRKKD 179

OY 171 RRRRKKD 178

DB 172 RRRRKKD 179

OY 171 RRRRKKD 178

DB 172 RRRRKKD 179

OY 171 RRRRKKD 178

DB 172 RRRRKKD 179

OY 171 RRRRKKD 178

DB 172 RRRRKKD 179

OY 171 RRRRKKD 178

AC 09PV23; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE XFD-11 PROTEIN.  
 GN XFD-11.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 NC NCBL\_TaxID=8355;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP Koechel W.;  
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Koester M., Dillinger K., Koechel W.;  
 RT "Expression pattern of the winged helix factor XFD-11 during Xenopus  
 embryogenesis.";  
 RL Mech. Dev. 76:169-173(1998).  
 DR EMBL; AJ242675; CAB44727.1;  
 DR HSSP; Q63245; 2HFH.  
 DR INTERPRO; IPR001766;  
 DR PFM; PF00250; Fork\_head. 1.  
 DR PRINTS; PR00053; FORKHEAD.  
 DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE; PS50039; FORK\_HEAD\_3; 1.  
 SQ SEQUENCE 492 AA; 53674 MW; 66906A5A9A2F04E CRC64;

Query Match 12.3%; Score 68; DB 13; Length 492;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-52; Indels 0; Gaps 0;  
 Matches 68; Conservative 0; Mismatches 0;

OY 111 RFFYDNNQGNQNSIRHLSLNECFVKVPRDDKKPGKGSYWTLPDPSYNNFENGSLRR 170  
 DB 112 RFFYDNNQGNQNSIRHLSLNECFVKVPRDDKKPGKGSYWTLPDPSYNNFENGSLRR 171  
 OY 171 RRRFRKKD 178  
 DB 172 RRRFRKKD 179

RESULT 6  
 ID P79771 PRELIMINARY; PRT; 445 AA.  
 AC P79771;  
 DT 01-MAY-1997 (TREMBlrel. 03, Created)  
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE WINGED HELIX PROTEIN CMH-2.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NC NCBL\_TaxID=9031;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE-EMBRYO.  
 RC MEDLINE-97141794; PubMed-8988052;  
 RA Freyaldenhoven B.S., Freyaldenhoven M.P., Iacovoni J.S., Vogt P.K.;  
 RT "Aberrant cell growth induced by avian winged helix proteins.";  
 RL Cancer Res. 57:123-129(1997).  
 DR EMBL; U37273; AAC60065.1;  
 DR HSSP; Q63245; 2HFH.  
 DR INTERPRO; IPR001766;  
 DR PFM; PF00250; Fork\_head. 1.  
 DR PRINTS; PR00053; FORKHEAD.  
 DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE; PS50039; FORK\_HEAD\_3; 1.

SQ SEQUENCE 445 AA; 48067 MW; 5F32D5CE521113F CRC64;

Query Match 11.0%; Score 61; DB 13; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 2.4e-46;  
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 NKQGMQNSIRHLSLNECFVKVPRDDKKPGKGSYWTLPDPSYNNFENGSLRRRRFRKK 177  
 DB 79 NKQGMQNSIRHLSLNECFVKVPRDDKKPGKGSYWTLPDPSYNNFENGSLRRRRFRKK 138  
 OY 178 D 178  
 DB 139 D 139

RESULT 7  
 ID O13017 PRELIMINARY; PRT; 468 AA.  
 AC O13017;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE WINGED HELIX TRANSCRIPTIONAL FACTOR MRF-1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NC NCBL\_TaxID=9031;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RA Arase Y., Koseki H., Miura N.;  
 RT Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.  
 RL EMBL; U95823; AAB53749.1;  
 DR HSSP; Q63245; 2HFH.  
 DR INTERPRO; IPR001766;  
 DR PFM; PF00250; Fork\_head. 1.  
 DR PRINTS; PR00053; FORKHEAD.  
 DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE; PS50039; FORK\_HEAD\_3; 1.  
 SQ SEQUENCE 468 AA; 50667 MW; EAFCA57ADF311746 CRC64;

Query Match 11.0%; Score 61; DB 13; Length 468;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-46;  
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 118 NKQGMQNSIRHLSLNECFVKVPRDDKKPGKGSYWTLPDPSYNNFENGSLRRRRFRKK 177  
 DB 112 NKQGMQNSIRHLSLNECFVKVPRDDKKPGKGSYWTLPDPSYNNFENGSLRRRRFRKK 171  
 OY 178 D 178  
 DB 172 D 172

RESULT 8  
 ID O9PVY8 PRELIMINARY; PRT; 461 AA.  
 AC O9PVY8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE FD-4' PROTEIN.  
 GN Xenopus laevis (African clawed frog).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 NC NCBL\_TaxID=8355;  
 OX [1]  
 RN SEQUENCE FROM N.A.

RA Koster M., Dillinger K., Knochel W.;  
 RT "Xenopus winged helix factors XFD-4/4" represent orthologues to  
 RL mammalian MFH-1.";  
 DR Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ249225; CAB54144.1;  
 DR HSSP: 063245; 2HFH.  
 DR INTERPRO: IPR001766;  
 DR PFAM: PF00250; Fork\_head; 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 SQ SEQUENCE 461 AA; 51162 MW; DE672C9F715BCF71 CRC64;

Query Match  
 Best Local Similarity 7.2%; Score 40; DB 13; Length 461;  
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 VKPPSYIALITMAIONAPDKKITTNGIYOFIMDRPFYR 116  
 DB 71 VKPPSYIALITMAIONAPDKKITTNGIYOFIMDRPFYR 110  
 |||||

RESULT 9  
 ID Q9PVY9 PRELIMINARY; PRT; 465 AA.  
 AC Q9PVY9;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 GN FD-4 PROTEIN.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OX NCBI\_TaxID=8353;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Koster M., Dillinger K., Knochel W.;  
 RT "Xenopus winged helix factors XFD-4/4" represent orthologues to  
 RL mammalian MFH-1.";  
 DR Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ249224; CAB54143.1;  
 DR HSSP: 063245; 2HFH.  
 DR INTERPRO: IPR001766;  
 DR PFAM: PF00250; Fork\_head; 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 SQ SEQUENCE 465 AA; 51613 MW; 15DFE9AD881E7245 CRC64;

Query Match  
 Best Local Similarity 7.2%; Score 40; DB 13; Length 465;  
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 VKPPSYIALITMAIONAPDKKITTNGIYOFIMDRPFYR 116  
 DB 71 VKPPSYIALITMAIONAPDKKITTNGIYOFIMDRPFYR 110  
 |||||

RESULT 10  
 ID Q9UDDO PRELIMINARY; PRT; 108 AA.  
 AC Q9UDDO;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DR FORKHEAD TRANSCRIPTIONAL ACTIVATOR HOMOLOG (FRAGMENT).  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-93271467; PubMed=849623;  
 RA Thomas R., Moore J., Johnston T., Socha C., Klemz M.;  
 RT "Drosophila forkhead homologues are expressed in a lineage-restricted  
 RL manner in human hematopoietic cells.";  
 DR Blood 81:2854-2859(1993).  
 DR HSSP: 063245; 2HFH.  
 DR INTERPRO: IPR001766;  
 DR PFAM: PF00250; Fork\_head; 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 SQ SEQUENCE 108 AA; 12886 MW; A6A1511F7D2899B2 CRC64;

Query Match  
 Best Local Similarity 7.1%; Score 39; DB 4; Length 108;  
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 113 PFYRDNKGQWNSIRHNLSTNECFVAVPRDCKPKGKSY 151  
 DB 44 PFYRDNKGQWNSIRHNLSTNECFVAVPRDCKPKGKSY 82  
 |||||

RESULT 11  
 ID Q22510 PRELIMINARY; PRT; 270 AA.  
 AC Q22510;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 GN T14G12.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 CC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Jones M., Kershaw J., Kirsten J., Hillier L., Jier M., Johnston L.,  
 RA Lightning J., Lloyd C., Murray A., Lalster N., Latreille P.,  
 RA Parsons J., Percy C., Rinken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RL elegans.";  
 RN Nature 368:32-38(1994).  
 RP SEQUENCE FROM N.A.  
 RA Wilcox L.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U41268; AA82436.1;  
 DR HSSP: 063245; 2HFH.  
 DR INTERPRO: IPR001766;  
 DR PFAM: PF00250; Fork\_head; 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS50039; FORK\_HEAD\_3; 1.



RA Treier M., Gleiberman A.S., O'Connell S.M., Szeto D.P., McMahon J.A.,  
 RA McMahon A.P., Rosenfeld M.G.,  
 RT "Multistep signaling requirements for pituitary organogenesis in  
 RT vivo.",  
 RL Genes Dev. 12:1691-1704(1998).  
 DR EMBL: AF060873; AAC27508.1; -.  
 DR HSSP: Q63245; 2HFH.  
 DR MGD: MGI:1349428; PERK.  
 DR INTERPRO: IPR001766; -.  
 DR PFAM: PF00250; Fork\_head; 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 FT NON\_TER 1 1  
 FT TER 90 90  
 SO SEQUENCE 90 AA; 10484 MW; 03CEDDEC6BA84BEB CRC64;

Query Match 2.9%; Score 16; DB 11; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GMONSTRHNLSNCF 136  
 Db 44 GMONSTRHNLSNCF 59

RESULT 15  
 ID Q9VZK0 PRELIMINARY; PRT: 365 AA.  
 AC Q9VZK0;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE FD64A PROTEIN.  
 GN FD64A.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_taxid=7227;  
 RX 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chape M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Butts K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Meyers A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz S., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svilaras R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Williams S.M., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,  
 RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AF003480; AAF47821.1; -.  
 DR HSSP: Q63245; 2HFH.  
 DR FLYBASE: FBgn0004895; fd64A.  
 DR INTERPRO: IPR001766; -.  
 DR PFAM: PF00250; Fork\_head; 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 SO SEQUENCE 365 AA; 40635 MW; 4FCE959A285FCA0F CRC64;

Query Match 2.9%; Score 16; DB 5; Length 365;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-06;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

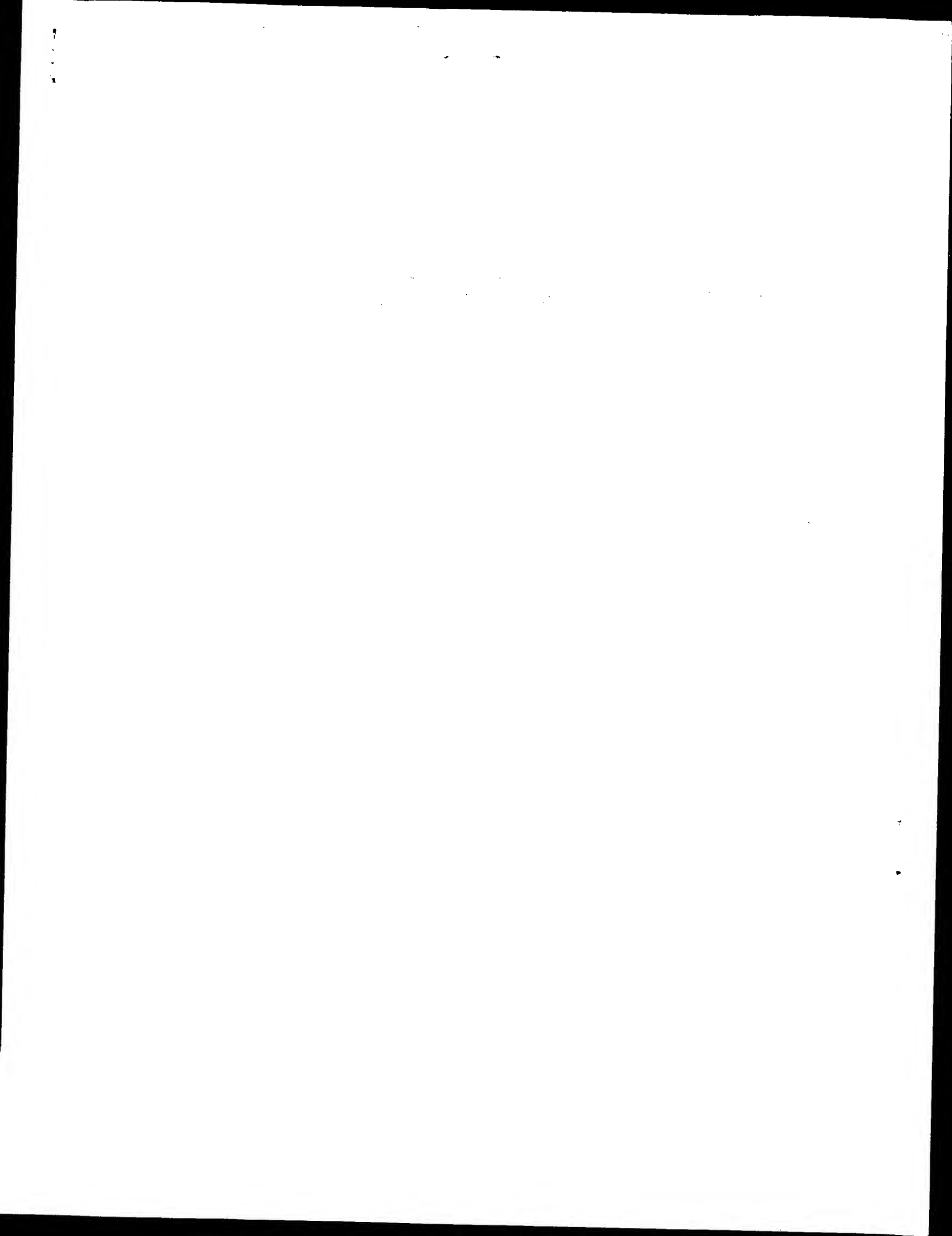
QY 118 NKGWONSTRHNLSN 133  
 Db 131 NKGWONSTRHNLSN 146

Search completed: February 6, 2001, 23:38:53  
 Job time: 6509 sec

Wed Feb 7 14:06:00 2001

us-09-292-862-2.rspt

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 6, 2001, 20:37:49 ; Search time 43.35 Seconds  
(Without alignments)  
229.071 Million cell updates/sec

Title: US-09-292-862-2  
Perfect score: 553  
Sequence: 1 MQARVSVSPNSLGVVPLG.....PSSQLRRTSGAFVYDCSKF 553

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 174772 seqs, 17957048 residues

Word size: 0 174772

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Issued Patents, AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/5.COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/PCITUS.COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	250	45.2	553	3 US-09-083-351-2	Sequence 2, Appl1
2	106	19.2	106	3 US-09-083-351-5	Sequence 5, Appl1
3	61	11.0	106	3 US-09-083-351-6	Sequence 6, Appl1
4	23	4.2	106	3 US-09-083-351-7	Sequence 7, Appl1
5	16	2.9	106	3 US-09-083-351-8	Sequence 8, Appl1
6	16	2.9	106	3 US-09-083-351-16	Sequence 16, Appl1
7	16	2.9	106	3 US-09-083-351-17	Sequence 17, Appl1
8	16	2.9	480	1 US-07-882-292-2	Sequence 2, Appl1
9	16	2.9	480	2 US-08-331-644-2	Sequence 2, Appl1
10	16	2.9	480	4 PCT-US93-04102-2	Sequence 12, Appl1
11	15	2.7	106	3 US-09-083-351-12	Sequence 13, Appl1
12	15	2.7	106	3 US-09-083-351-13	Sequence 13, Appl1
13	13	2.4	76	3 US-09-083-351-19	Sequence 19, Appl1
14	13	2.4	98	1 US-07-906-930E-4	Sequence 4, Appl1
15	13	2.4	106	3 US-09-083-351-11	Sequence 11, Appl1
16	13	2.4	540	1 US-07-906-930E-2	Sequence 2, Appl1
17	13	2.2	106	3 US-09-083-351-20	Sequence 20, Appl1
18	11	2.0	37	1 US-08-425-069-35	Sequence 35, Appl1
19	11	2.0	37	1 US-08-425-069-36	Sequence 46, Appl1
20	11	2.0	37	1 US-08-209-747-46	Sequence 47, Appl1
21	11	2.0	37	1 US-08-209-747-47	Sequence 47, Appl1
22	11	2.0	37	1 US-08-458-298-46	Sequence 46, Appl1
23	11	2.0	37	1 US-08-458-298-47	Sequence 47, Appl1
24	11	2.0	37	2 US-08-317-844B-35	Sequence 35, Appl1
25	11	2.0	37	2 US-08-317-844B-36	Sequence 36, Appl1
26	11	2.0	595	2 US-08-425-069-4	Sequence 4, Appl1
27	11	2.0	595	2 US-08-317-844B-4	Sequence 4, Appl1
28	11	2.0	4472	2 US-08-804-227C-2	Sequence 2, Appl1

29	11	2.0	4545	2 US-08-804-227C-14	Sequence 14, Appl1
30	10	1.8	10	1 US-07-969-307A-20	Sequence 20, Appl1
31	10	1.8	18	2 US-08-528-129A-10	Sequence 10, Appl1
32	10	1.8	20	1 US-08-234-602-14	Sequence 14, Appl1
33	10	1.8	35	1 US-08-209-747-52	Sequence 52, Appl1
34	10	1.8	35	1 US-08-458-298-52	Sequence 52, Appl1
35	10	1.8	98	1 US-07-906-930E-6	Sequence 9, Appl1
36	10	1.8	106	3 US-09-083-351-9	Sequence 9, Appl1
37	10	1.8	106	3 US-09-083-351-10	Sequence 10, Appl1
38	10	1.8	106	3 US-09-083-351-14	Sequence 14, Appl1
39	10	1.8	106	3 US-09-083-351-15	Sequence 15, Appl1
40	10	1.8	148	1 US-08-207-904-15	Sequence 15, Appl1
41	10	1.8	161	2 US-08-581-528A-6	Sequence 6, Appl1
42	10	1.8	161	4 PCT-US94-07799-6	Sequence 6, Appl1
43	10	1.8	199	3 US-08-506-553C-8	Sequence 8, Appl1
44	10	1.8	208	2 US-08-935-886-8	Sequence 8, Appl1
45	10	1.8	211	2 US-08-935-886-10	Sequence 10, Appl1

## ALIGNMENTS

RESULT 1  
US-09-083-351-2  
Sequence 2, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
TITLE OF INVENTION: TRANSCRIPTION FACTOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: FOLEY, HONG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UTA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEO ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 553 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-2

Query Match 45.2%; Score 250; DB 3; Length 553;  
Best Local Similarity 99.7%; Pred. No. 2.7e-215;  
Matches 350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 203 PAPEQADGNAPGPPVPRIDIKTENGTCSPFPLSPAAALSGSAAVPKIESPDS 262

```

|||||
Db 203 PAPPEADGNAGPQPPPPRIQDIXTEGTCSPPPPLSPAALGSSAAAVPKIESPDS 262
QY 263 SSSSSSSGSPSPSLPSARPLSLDGDADSAPPPAPSAAPPHHSGGFSVDNIMTSLRSPQ 322
Db 263 SSSSSSSGSPSPSLPSARPLSLDGDADSAPPPAPSAAPPHHSGGFSVDNIMTSLRSPQ 322
QY 323 SAAAEELSSGLLASAAASSRAGIAPPLALGAVSPGSSSYSPGSGTSSASSSGGGGAG 382
Db 323 SAAAEELSSGLLASAAASSRAGIAPPLALGAVSPGSSSYSPGSGTSSASSSGGGGAG 382
QY 383 AAGGAGAGTTCNLTQMSLYAAGERGHLOGAPGAGSADVDPEDVSLPVTSSSSS 442
Db 383 AAGGAGAGTTCNLTQMSLYAAGERGHLOGAPGAGSADVDPEDVSLPVTSSSSS 442
QY 443 SLSHGGGGGGGGGGGQHHPAHOGRITSMYLNQAGGDLGHTLASAAAAAAGYGGQQ 502
Db 443 SLSHGGGGGGGGGGGQHHPAHOGRITSMYLNQAGGDLGHTLASAAAAAAGYGGQQ 502
QY 503 NFHSVREMFESQRTGLNNSPVNGSSCOMAFSSQSLYRTSGAFVYDCKF 553
Db 503 NFHSVREMFESQRTGLNNSPVNGSSCOMAFSSQSLYRTSGAFVYDCKF 553

```

```

RESULT 2
US-09-083-351-5
; Sequence 5, Application US/09083351
; Patent No. 6087107
GENERAL INFORMATION:
APPLICANT: Sheffield, Val C.
APPLICANT: Alward, Wallace L.M.
APPLICANT: Stone, Edwin M.
APPLICANT: Nishimura, Darryl
APPLICANT: Patil, Shiva
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,351
FILING DATE: 22-MAY-1998
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-029.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 106 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-083-351-5

```

Query Match 19.2%; Score 106; DB 3; Length 106;  
Best Local Similarity 100.0%; Pred. No. 2.3e-87;

```

Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 73 PKDWMKPPYSTALITMAIONAPDKKITLNGIYQFIMDRFPFYNKOGMONSTRINLSI 132
Db 1 PKDWMKPPYSTALITMAIONAPDKKITLNGIYQFIMDRFPFYNKOGMONSTRINLSI 60
QY 133 NECFYKVRDCKKPKGKSYWTLDPDSYNMFENGSLFRRRRFRKKD 178
Db 61 NECFYKVRDCKKPKGKSYWTLDPDSYNMFENGSLFRRRRFRKKD 106

```

```

RESULT 3
US-09-083-351-6
; Sequence 6, Application US/09083351
; Patent No. 6087107
GENERAL INFORMATION:
APPLICANT: Sheffield, Val C.
APPLICANT: Alward, Wallace L.M.
APPLICANT: Stone, Edwin M.
APPLICANT: Nishimura, Darryl
APPLICANT: Patil, Shiva
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,351
FILING DATE: 22-MAY-1998
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UIA-029.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 106 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-083-351-6

```

Query Match 11.0%; Score 61; DB 3; Length 106;  
Best Local Similarity 100.0%; Pred. No. 3.6e-47;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 118 NKOGWMSIRHNLSINECFYKVRDCKKPKGKSYWTLDPDSYNMFENGSLFRRRRFRKK 177  
Db 46 NKOGWMSIRHNLSINECFYKVRDCKKPKGKSYWTLDPDSYNMFENGSLFRRRRFRKK 105  
QY 178 D 178  
Db 106 D 106  
RESULT 4  
US-09-083-351-7

Sequence 7, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
TITLE OF INVENTION: TRANSCRIPTION FACTOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-7

Query Match 4.28: Score 23: DB 3: Length 106;  
Best Local Similarity 100.0%; Pred. No. 3.1e-13;  
Matches 23: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GWNSTIRNLSINCEFCVAVPRDD 143  
|||||  
DB 49 GWNSTIRNLSINCEFCVAVPRDD 71

RESULT 5  
US-09-083-351-8  
Sequence 8, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
TITLE OF INVENTION: TRANSCRIPTION FACTOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA

COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-8

Query Match 2.98: Score 16: DB 3: Length 106;  
Best Local Similarity 100.0%; Pred. No. 5.6e-07;  
Matches 16: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 TLNGTIOFIMDRPEFY 115  
|||||  
DB 28 TLNGTIOFIMDRPEFY 43

RESULT 6  
US-09-083-351-16  
Sequence 16, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
TITLE OF INVENTION: TRANSCRIPTION FACTOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000

TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-16

Query Match 2.9%; Score 16; DB 3; Length 106;  
Best Local Similarity 100.0%; Pred. No. 5.6e-07;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKOGWONSIRHNLSLN 133  
DB 46 NKOGWONSIRHNLSLN 61

RESULT 7  
US-09-083-351-17  
Sequence 17, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Dairyl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESS: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-17

Query Match 2.9%; Score 16; DB 3; Length 106;  
Best Local Similarity 100.0%; Pred. No. 5.6e-07;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 118 NKOGWONSIRHNLSLN 133  
DB 46 NKOGWONSIRHNLSLN 61

DB 46 NKOGWONSIRHNLSLN 61

RESULT 8  
US-07-882-292-2  
Sequence 2, Application US/07882292  
Patent No. 5324638  
GENERAL INFORMATION:  
APPLICANT: Tao, Wufan  
APPLICANT: Lai, Eseng  
TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC ACIDS  
TITLE OF INVENTION: ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESS: John P. White  
STREET: c/o Cooper and Dunham, 30 Rockefeller  
STREET: Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/882,292  
FILING DATE: 19920513  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41472  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-977-9550  
TELEFAX: 212-664-0525  
TELLEX: 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 480 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-882-292-2

Query Match 2.9%; Score 16; DB 1; Length 480;  
Best Local Similarity 100.0%; Pred. No. 2.3e-06;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKOGWONSIRHNLSLN 133  
DB 212 NKOGWONSIRHNLSLN 227

RESULT 9  
US-08-331-644-2  
Sequence 2, Application US/08331644  
Patent No. 5976872  
GENERAL INFORMATION:  
APPLICANT: Tao, Wufan  
APPLICANT: Lai, Eseng  
TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC  
TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESS: Cooper & Dunham  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/331,644  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,292  
FILING DATE: 13-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41472-A-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-391-0525  
TELEFAX: 212-278-0400  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 480 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-331-644-2

Query Match  
Best Local Similarity 100.0%; Score 16; DB 2; Length 480;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKQGMONSIRHNSLN 133  
DB 212 NKQGMONSIRHNSLN 227

RESULT 10  
PCT-US93-04102-2  
Sequence 2, Application PC/TUS9304102  
GENERAL INFORMATION:  
APPLICANT: Tao, Wufan  
APPLICANT: Lai, Eseng  
TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NUCLEIC  
TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04102  
FILING DATE: 19930430  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,292  
FILING DATE: 13-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 41472A-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-664-0525  
TELEFAX: 422523 COOP UI

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 480 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-04102-2

Query Match  
Best Local Similarity 100.0%; Score 16; DB 4; Length 480;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 NKQGMONSIRHNSLN 133  
DB 212 NKQGMONSIRHNSLN 227

RESULT 11  
US-09-083-351-12  
Sequence 12, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
TITLE OF INVENTION: TRANSCRIPTION FACTOR  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HONG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-083-351-12

Query Match  
Best Local Similarity 100.0%; Score 15; DB 3; Length 106;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 VKPPSYATLITMAI 91  
DB 5 VKPPSYATLITMAI 19

RESULT 12  
US-09-083-351-13  
; Sequence 13, Application US/09083351  
; Patent No. 6087107  
; GENERAL INFORMATION:  
; APPLICANT: Sheffield, Val C.  
; APPLICANT: Alward, Wallace L.M.  
; APPLICANT: Stone, Edwin M.  
; APPLICANT: Nishimura, Darryl  
; APPLICANT: Patil, Shiva  
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/083,351  
; FILING DATE: 22-MAY-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arnold, Beth E.  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: UIA-029.02  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-1000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ. ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 106 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-083-351-13

Query Match 2.7%; Score 15; DB 3; Length 106;  
Best Local Similarity 100.0%; Pred. No. 4.3e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 VKPPSYIALITMAI 91  
Db 5 VKPPSYIALITMAI 19

RESULT 13  
US-09-083-351-19  
; Sequence 19, Application US/09083351  
; Patent No. 6087107  
; GENERAL INFORMATION:  
; APPLICANT: Sheffield, Val C.  
; APPLICANT: Alward, Wallace L.M.  
; APPLICANT: Stone, Edwin M.  
; APPLICANT: Nishimura, Darryl  
; APPLICANT: Patil, Shiva  
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
; TITLE OF INVENTION: TRANSCRIPTION FACTOR  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP

STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-7000  
; INFORMATION FOR SEQ. ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 76 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-083-351-19

Query Match 2.4%; Score 13; DB 3; Length 76;  
Best Local Similarity 100.0%; Pred. No. 0.00019;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 GWNQSIHNLNLSIN 133  
Db 49 GWNQSIHNLNLSIN 61

RESULT 14  
US-07-906-930E-4  
; Sequence 4, Application US/07906930E  
; Patent No. 5534631  
; GENERAL INFORMATION:  
; APPLICANT: Gaynor, Richard B.  
; APPLICANT: Nirula, Ajay  
; APPLICANT: Li, Ching  
; TITLE OF INVENTION: DNA ENCODING THE INTERLEUKIN BINDING  
; TITLE OF INVENTION: FACTOR (ILF)  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/906,930E  
; FILING DATE: 30-JUN-1992  
; CLASSIFICATION: 336  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sertlich, Gary J.  
; REGISTRATION NUMBER: 34,430  
; REFERENCE/DOCKET NUMBER: UTSD:262/SER  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 512-418-3000

TELEFAX: 512-474-7577  
 TELEX: NOT APPLICABLE  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 98 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-07-906-930E-4

Query Match 2.4%; Score 13; DB 1; Length 98;  
 Best Local Similarity 100.0%; Pred. No. 0.00025;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 121 GWONSIRHNLSLN 133  
 ||||||||||||  
 Db 46 GWONSIRHNLSLN 58

RESULT 15

US-09-083-351-11  
 ; Sequence 11, Application US/09083351  
 ; Patent No. 6087107  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sheffield, Val C.  
 ; APPLICANT: Alward, Wallace L.M.  
 ; APPLICANT: Stone, Edwin M.  
 ; APPLICANT: Nishimura, Darryl  
 ; APPLICANT: Patil, Shiva  
 ; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
 ; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
 ; TITLE OF INVENTION: TRANSCRIPTION FACTOR  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: FOLEY, HONG & ELIOT LLP  
 ; STREET: One Post Office Square  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02109-2170  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/083,351  
 ; FILING DATE: 22-MAY-1998  
 ; CLASSIFICATION: 514  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Arnold, Beth E.  
 ; REGISTRATION NUMBER: 35,430  
 ; REFERENCE/DOCKET NUMBER: UIA-029.02  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-832-1000  
 ; TELEFAX: 617-832-7000  
 ; INFORMATION FOR SEQ ID NO: 11:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 106 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-09-083-351-11

Query Match 2.4%; Score 13; DB 3; Length 106;  
 Best Local Similarity 100.0%; Pred. No. 0.00027;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 121 GWONSIRHNLSLN 133

Db 49 GWONSIRHNLSLN 61  
 ||||||||||||

Search completed: February 6, 2001, 22:53:53  
 Job time: 8164 sec





Wed Feb 7 14:05:55 2001

us-09-292-862-1.rst

Page 1

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 6, 2001, 18:03:13 ; Search time 966.16 Seconds  
(without alignments)  
12054.364 Million cell updates/sec

Title: US-09-292-862-1  
Perfect score: 1662  
Sequence: 1 atgcagcgcgctactccgt.....acgactgtagcaagtlttga 1662

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 80.0

Searched: 7991742 seqs, 3503743858 residues

Word size : 0 15983484

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*\*  
2: gb\_est2:\*\*  
3: gb\_est3:\*\*  
4: gb\_est4:\*\*  
5: gb\_est5:\*\*  
6: gb\_est6:\*\*  
7: gb\_est7:\*\*  
8: gb\_est8:\*\*  
9: gb\_est9:\*\*  
10: gb\_est10:\*\*  
11: gb\_est11:\*\*  
12: gb\_est12:\*\*  
13: gb\_est13:\*\*  
14: gb\_est14:\*\*  
15: gb\_est15:\*\*  
16: gb\_est16:\*\*  
17: gb\_est17:\*\*  
18: gb\_est18:\*\*  
19: gb\_est19:\*\*  
20: gb\_est20:\*\*  
21: gb\_est21:\*\*  
22: gb\_est22:\*\*  
23: gb\_est23:\*\*  
24: gb\_est24:\*\*  
25: gb\_est25:\*\*  
26: gb\_est26:\*\*  
27: gb\_est27:\*\*  
28: gb\_est28:\*\*  
29: gb\_est29:\*\*  
30: gb\_est30:\*\*  
31: gb\_est31:\*\*  
32: gb\_est32:\*\*  
33: gb\_est33:\*\*  
34: gb\_est34:\*\*  
35: gb\_est35:\*\*  
36: gb\_est36:\*\*  
37: gb\_est37:\*\*  
38: gb\_est38:\*\*  
39: gb\_est39:\*\*  
40: gb\_est40:\*\*  
41: em\_estba:\*\*  
42: em\_estfun:\*\*  
43: em\_esthum1:\*\*

44: em\_esthum2:\*\*  
45: em\_esthum3:\*\*  
46: em\_esthum4:\*\*  
47: em\_esthum5:\*\*  
48: em\_esthum6:\*\*  
49: em\_esthum7:\*\*  
50: em\_esthum8:\*\*  
51: em\_esthum9:\*\*  
52: em\_esthum10:\*\*  
53: em\_esthum11:\*\*  
54: em\_esthum12:\*\*  
55: em\_esthum13:\*\*  
56: em\_esthum14:\*\*  
57: em\_esthum15:\*\*  
58: em\_esthum16:\*\*  
59: em\_esthum17:\*\*  
60: em\_esthum18:\*\*  
61: em\_esthum19:\*\*  
62: em\_esthum20:\*\*  
63: em\_estin1:\*\*  
64: em\_estin2:\*\*  
65: em\_estin3:\*\*  
66: em\_estin4:\*\*  
67: em\_estov1:\*\*  
68: em\_estov2:\*\*  
69: em\_estpl1:\*\*  
70: em\_estpl2:\*\*  
71: em\_estpl3:\*\*  
72: em\_estpl4:\*\*  
73: em\_estpl5:\*\*  
74: em\_estro1:\*\*  
75: em\_estro2:\*\*  
76: em\_estro3:\*\*  
77: em\_estro4:\*\*  
78: em\_estro5:\*\*  
79: em\_estro6:\*\*  
80: em\_estro7:\*\*  
81: em\_estro8:\*\*  
82: em\_estro9:\*\*  
83: em\_estro10:\*\*  
84: em\_estro11:\*\*  
85: em\_estro12:\*\*  
86: em\_estro13:\*\*  
87: gb\_esta1:\*\*  
88: gb\_esta2:\*\*  
89: gb\_esta3:\*\*  
90: gb\_esta4:\*\*  
91: gb\_esta5:\*\*  
92: gb\_esta6:\*\*  
93: gb\_esta7:\*\*  
94: gb\_esta8:\*\*  
95: gb\_esta9:\*\*  
96: gb\_esta10:\*\*  
97: gb\_esta11:\*\*  
98: gb\_esta12:\*\*  
99: gb\_esta13:\*\*  
100: gb\_esta14:\*\*  
101: gb\_esta15:\*\*  
102: gb\_esta16:\*\*  
103: gb\_esta17:\*\*  
104: gb\_esta18:\*\*  
105: gb\_esta19:\*\*  
106: gb\_esta20:\*\*  
107: gb\_esta21:\*\*  
108: gb\_esta22:\*\*  
109: gb\_esta23:\*\*  
110: gb\_esta24:\*\*  
111: gb\_esta25:\*\*  
112: em\_esthum21:\*\*  
113: em\_esthum22:\*\*  
114: em\_esthum23:\*\*  
115: em\_estom1:\*\*  
116: em\_estom2:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
117: em_estp16:*	305	18.4	305	23	A1659865	A1659865 tu18h10.x
118: em_estp17:*	263	15.8	555	24	A1761460	A1761460 w65h10.x
119: em_estp18:*	249	15.0	305	25	A1817619	A1817619 w39f10.x
120: em_estp19:*	231	13.9	507	19	A1333220	A1333220 q905g08.x
121: em_estp20:*	186	11.2	504	25	A1819982	A1819982 w760a05.x
122: em_estp21:*	185	11.1	259	90	AW451944	AW451944 UI-H-B13-
123: em_estp22:*	185	11.1	292	90	AW451610	AW451610 UI-H-B13-
124: em_estp23:*	169	10.2	439	1	AA022618	AA022618 ze71a01.s
125: em_estp24:*	163	9.8	223	25	A1804844	A1804844 t195a09.x
126: em_estp25:*	156	9.4	180	24	A1763387	A1763387 w155h12.x
127: em_estp26:*	155	9.3	579	87	AW195359	AW195359 x037d07.x
128: em_estp27:*	150	9.0	242	88	AW263103	AW263103 x035h09.x
129: em_estp28:*	138	8.3	138	148	AF190834	AF190834 AF190834
130: em_estp29:*	135	8.1	486	110	BE644653	BE644653 t768f05.x
131: em_estp30:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
132: em_estp31:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
133: em_estp32:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
134: em_estp33:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
135: em_estp34:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
136: em_estp35:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
137: em_estp36:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
138: em_estp37:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
139: em_estp38:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
140: em_estp39:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
141: em_estp40:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
142: em_estp41:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
143: em_estp42:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
144: em_estp43:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
145: em_estp44:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
146: em_estp45:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
147: em_estp46:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
148: em_estp47:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
149: em_estp48:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
150: em_estp49:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
151: em_estp50:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
152: em_estp51:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
153: em_estp52:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
154: em_estp53:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
155: em_estp54:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
156: em_estp55:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
157: em_estp56:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
158: em_estp57:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
159: em_estp58:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
160: em_estp59:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
161: em_estp60:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
162: em_estp61:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
163: em_estp62:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
164: em_estp63:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
165: em_estp64:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
166: em_estp65:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
167: em_estp66:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
168: em_estp67:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
169: em_estp68:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
170: em_estp69:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
171: em_estp70:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
172: em_estp71:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
173: em_estp72:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
174: em_estp73:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
175: em_estp74:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
176: em_estp75:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
177: em_estp76:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
178: em_estp77:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
179: em_estp78:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
180: em_estp79:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
181: em_estp80:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
182: em_estp81:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
183: em_estp82:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
184: em_estp83:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
185: em_estp84:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
186: em_estp85:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
187: em_estp86:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
188: em_estp87:*	135	8.1	436	20	A1453830	A1453830 t747e05.x
189: em_estp88:*	135	8.1	436	20	A1453830	A1453830 t747e05.x

## ALIGNMENTS

RESULT 1  
LOCUS A1659865/c  
DEFINITION tu18h10.x1 NCI-CCAP\_P128 Homo sapiens CDNA clone image-2251459 3'  
; contains element PTR5 repetitive element ; mRNA sequence.  
ACCESSION A1659865  
VERSION A1659865.1 GI:4763435  
KEYWORDS EST.





[illegible]

Db	1	GCGCCGACAGCTACCCGGGCCAGCAGCAACTTCCACTCGCTCGGAGATGTTCCAG	60
Qy	1537	tcacagagatcgcgtctgtaacaacatctccagtgaacgggaatagatgttcaatggcc	1596
Db	61	TCACAGAGGATCGCGCTTGAGACACTCTCCAGTGAACGGGAATGATGCTCAATATGGCC	120
Qy	1597	ttcccttcacgaagctctctgtatccgcacagctccgcagagcttcgctctacgactgtagcaag	1656
Db	121	TTCCCTTCACGACAGTCTGTATCCGACGCTCCGAGCTTTCTGCTACAGACTGTAGCAAG	180
Qy	1657	ttttga 1662	
Db	181	TTTTGA 186	

RESULT	6		EST	17-FEB-2000
AW451944	259 bp	mrna		
LOCUS	UI-H-R13-alt-g-08-0-ui.s1	NCI_CGAP_Sub5	Homo sapiens	CDNA clone
DEFINITION	IMAGE3068727 3',	mrna sequence.		
ACCESSION	AW451944			
VERSION	AW451944.1	GI:6992720		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 259)			
JOURNAL	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .			
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index			
	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			
	Tel: (301) 496-1550			
	Email: Robert.L.Strausberg@nih.gov			
	Oligo-qr track not found, Not I site shown in beginning of sequence			
	is likely internal to the message. CDNA Library Preparation: M.B.			
	Soares Lab Clone distribution: NCI-CGAP clone distribution			
	information can be found through the I.M.A.G.E. Consortium/LLNL at:			
	<a href="http://www.bio.llnl.gov/dbrr/image/image.html">www.bio.llnl.gov/dbrr/image/image.html</a>			
	Seq primer: M13 Forward			
	POLYA=NO.			

FEATURES	source	Location/Qualifiers
	1..259	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:3068727"	
	/clone_lib="NCI_CGAP_Sub5"	
	/lab_host="DH10B (Life Technologies)"	
	/note="Vector: pT73D-Pac (Pharmacia) with a modified	
	polylinker; Site_1: Not I; Site_2: Eco RI. NCI_CGAP_Sub5	
	is a subtracted library derived from NCI_CGAP_Sub4. The	
	NCI_CGAP_Sub5 library had 3 million recombinants. A	
	single-stranded DNA preparation of NCI_CGAP_Sub4 was used	
	as a tracer in a subtractive hybridization with a driver	
	comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLM	
	3334-3337, 3682-3683, 3799-3803 (IMAGE Cloneds	
	1322376-1323911, 1456008-1456775, 1500552-1502855);	
	NCI_CGAP_Kid5 pool 1 LLM 3338-3342, 3722-3725, 3776-3778	
	(IMAGE Cloneds 1323912-1325831, 1471368-1472903,	
	1492104-1493255); NCI_CGAP_Lu5 pool 1 LLM 3575-3582,	
	3851-3854 (IMAGE Cloneds 1414920-1417991, 1520904-1522433	
	); NCI_CGAP_Gc4 pool 1 LLM 3164-3167, 3716-3720,	
	3733-3735 (IMAGE Cloneds 1257096-1258631, 1469064-1470983	
	1475592-1476743); NCI_CGAP_Pt22 pool 1 LLM 2457-2459,	
	2758-2759, 3062-3068 (IMAGE Cloneds 985608-986759	
	1101192-1101959, 1211928-1220615); NCI_CGAP_Co10 pool 1	
	LLM 2644-2653, 2871-2872 (IMAGE Cloneds 1057416-106125	
	, 114584-1145351). (10% of the driver population), plus a	
	pool of 3,840 arrayed clones from NCI_CGAP_Sub1 (IMAGE	
	Cloneds 2708616-2710535) and NCI_CGAP_Sub2 (IMAGE	
	Cloneds 2710536-2712455) (10% of the driver population	

	Query Match	11.1%	Score 185	DB 90	Length 259
	Best Local Similarity	100.0%	Pred. NO. 4.9e-76		
	Matches 185	Conservative	0	Mismatches 0	Indels 0
				Gaps 0	
QY	1478	cggcgccagagctaccgggcccagacaggaactccactcggatgcyggaatgcttcgagtc	1537		
Db	1	CGGCGCCAGAGCTACCGGGGCCACGACAGAACTTCACCTCGTCCGGGAGATGTTGCAGT	60		
QY	1538	caagaagatctggcttgaacaactctccagtgaacggygaatgagctgtccaatgacct	1597		
Db	61	CACAGAGATCTGGCTTGAACAACCTTCCAGTGAACGGGGAATAGTACCTGTCAAAATGGCTT	120		
QY	1598	ttcccttcagcagcagctctctgtatccgcaagcttcggagcttcgtctctaagacttgaagaat	1657		
Db	121	TCCTTCACGCCACGACTCTGTATCCGACACCTTCGGAGCTTCTGCTACGACTTACGAAAT	180		
QY	1658	ttctga 1662			
Db	181	TTTGA 185			

```

Location/Qualifiers
1..292
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2736695"
/clone_lbp="NCI_CGAP_Sub5"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT733-Pac (pharmacia) with a modified
polylinker. Site1: Not I, Site2: Eco RI; NCI_CGAP_Sub5
is a subtracted library derived from NCI_CGAP_Sub4."

```

Query Match	11.18;	Score 185;	DB 90;	Length 292;
Best Local Similarity	100.0%;	Pred. No. 4.9e-76;		
Matches 185;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1478	cgagccgacggtaccgagccgacgaagaacttccactcgttcgagagatggttcgaagt	1537		
Db 1	CGGCGCGAGGCTTACCCGGGCCACAGACAGACTTCCTACTCGGTGCGGAGATGTTCCAGT	60		
QY 1538	cacagagatcgcgtttgaacaacttcccaagtgaacgggaatagatgctgtcaaatggcct	1597		
Db 61	CACAGAGATGCGTTTGAACAACCTTCACAGTGAACGGGAATAATGACTGTCAAAATGGCC	120		
QY 1598	tcacctccagccagctcctctgttaccggaacgtccgagacttcgtttaccagatgttagaagt	1657		
Db 121	TCCCTTCACGACGAGTCTCTGTATCCGACGAGTCGGAGCTTTTGTTACGACTGTACGCAAGT	180		
QY 1658	tttga 1662			
Db 181	TTTGA 185			

RESULT 8	AA022618	LOCUS	AA022618	439 bp	mRNA	EST	30-JAN-1997
DEFINITION	ze71a01.s1 Soares-fetal heart NDH19M						Homo sapiens cDNA clone
ACCESSION	AA022618						mouse
VERSION	AA022618.1						
KEYWORDS	EST.						
SOURCE	human.						
ORGANISM	Homo sapiens						

ORGANISM	
<i>Homo sapiens</i>	Cranialia; Vertebrata; Euteleostomi;
<i>Eukaryota</i> ; Metazoa; Chordata	
<i>Mammalia</i> ; <i>Eutheria</i> ; Primates; Catarrhini; Hominoidea; Homo.	
1 (bases 1 to 439)	
REFERENCE	
AUTHORS	
Chiller, S., Lennon, G., Becker, M., Bonaldo, M. F., Chippelli, B., Hillier, S., Dietrich, N., Dubugue, T., Favell, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacey, M., Le, M., Le, N., Maritz, E., MOO	

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.









QY 1477 qcqgcgcacactaccggcgccgccccccc

Db 1 GGGGGCCGAGGCTACCCGGGGCCAGCAGCAAACTTCCACTCGGCGGGGAGATGTTCCAG 60  
 1537 tcacagagatcgcgcttgaaacaactctccagtgaaacgggaatagtagctgtcaaatgccc 1596  
 QY 61 ttacacagagatcgcgcttgaaacaactctccagtgaaacgggaatagtagctgtcaaatgccc 1596

QY	159/	ttcccttcacgacg	ctcgtaccgacg	tcggagcgttc	tgcttaagactg	atgacaag	165
Db	121	tttccttccacgacg <th>ctcgtaccgacg</th> <th>tcggagcgttc</th> <th>tgcttaagactg</th> <th>atgacaag</th> <th>180</th>	ctcgtaccgacg	tcggagcgttc	tgcttaagactg	atgacaag	180
QY	1657	ttttta	1662				
Db	181	TTTTGA	186				

RESULT 15  
AI453830

LOCUS	AI453830	436 bp	mRNA	EST	13-APR-1999
DEFINITION	tj47e05.x1	Soares NSF	FR 9W OT PA	D C1	

IMAGE:21446/2.3' similar to SW:MEF1\_HUMAN Q99958 MESENCHYME FORKHEAD PROTEIN 1 ;, mRNA sequence.  
AT152820

VERSION AI453830.1 CI:4285047  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
1 (bases 1 to 435)  
Mammalia; Eutheria;  
Primates; Catarrhini; Hominoidea; Homo.  
Chiroptera; Euteleostomi;  
Vertebrata; Euteleostomi;  
Carnivora; Carnivora; Canidae; Felidae; Mustelidae; Ursidae; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

TITLE  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP)  
 URL  
 NC-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D

Email: Robert\_Strausberg@nih.gov  
This clone is available for research purposes only.

Insert length: 551 Std Error: 0.00  
See primary - down from 551 to 550 for further information.

High quality sequence stop: 413.  
Location/Qualifiers

**source**

1.436

```

/db_xref="taxon:9606"

```

```
/c)one 1ib="sources you no ...
/c)one IMAGE:2144672"
```

```
/lab_host="DH10B"
```

a modified polylinker; Site 1: Not I; Site 2: Eco RI.

libraries were mixed, and ss circles were normalized

For hybridization, this DNA was used as tracer in

for-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries.

libraries and cloneIDs: Soares NbHSF pool 1:

145032-147335, 147720-148103, 148872-149255 15003  
codes NZHP pool 1:

758280-760583, 772104-774407 Soares Nbhuda pool 1:

320136-322823, 326280-326663 Soares NbHOT

pool 1: 723720-726407, 739080-740999	Subtraction by Bence
Soares and M. Fatima Bonaldo."	
BASE COUNT	156 a 119 c 75 g 86 t
ORIGIN	

Query Match	6.78;	Score 112;	DB 20;	Length 436;
Best Local Similarity	-99.48;	Pred. No. 7.8e-42;		
Matches 162;	Conservative 0;	Mismatches 1;	Indels 0;	Gap 0;

[illegible]

Db 61 CTGCCAGTGAAGGGAATATCTAGCTGTCAATGGGCTTCCCTTCCAGCCAGTCTCTGTA 120  
1620 ccgacagctccgagcttcgcgtcctgacgactcgtcgcgaagcttga 1662  
121 CCGACGCTCCGGAGCTTTCGTCTACACACGTGAGCAAGTTTGA 163

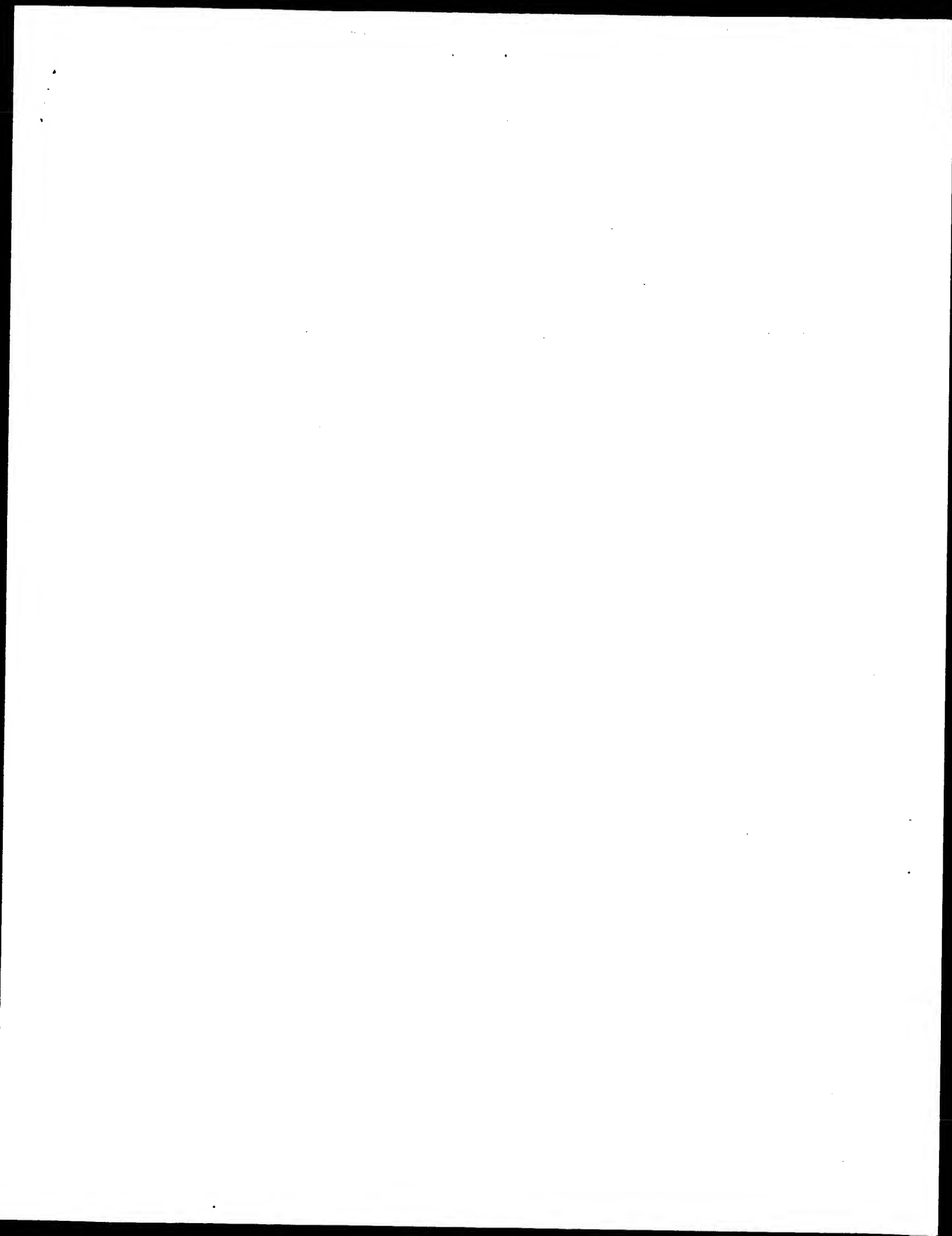
Search completed: February 6, 2001, 20:37:45  
Job time: 9272 sec

Wed Feb 7 14:05:55 2001

us-09-292-862-1.rst

---

Page 11



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 6, 2001, 18:24:39 : Search time 51.77 Seconds  
(without alignments)  
5173.815 Million cell updates/sec

Title: US-09-292-862-1

Perfect score: 1662  
Sequence: 1 atgcagcgccgctactcgt.....acgactgtagcaatttga 1662

Scoring table: OLIGO\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 280836 seqs, 80580151 residues

Word size: 0

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/PCtrus.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1217	73.2	3946	US-09-083-351-1	Sequence 1, Appli
2	1214	73.0	1659	US-09-083-351-3	Sequence 3, Appli
3	47	2.8	1155	US-08-331-644-4	Sequence 4, Appli
4	47	2.8	1155	US-08-331-644-3	Sequence 4, Appli
5	28	1.7	1860	US-08-331-644-3	Sequence 3, Appli
6	28	1.7	1860	US-08-331-644-3	Sequence 3, Appli
7	27	1.6	640	US-08-835-099A-16	Sequence 16, Appli
8	27	1.6	640	US-09-157-349-16	Sequence 16, Appli
9	27	1.6	804	US-08-835-099A-10	Sequence 10, Appli
10	27	1.6	804	US-09-157-349-10	Sequence 10, Appli
11	25	1.5	3330	US-08-149-103-1	Sequence 1, Appli
12	25	1.5	3330	US-08-451-883-1	Sequence 1, Appli
13	25	1.5	3656	US-08-393-734-1	Sequence 1, Appli
14	25	1.5	8438	US-07-945-283-1	Sequence 1, Appli
15	25	1.5	9592	US-08-393-734-3	Sequence 3, Appli
16	25	1.5	25	US-08-374-144-3	Sequence 3, Appli
17	24	1.4	25	US-08-775-164-3	Sequence 3, Appli
18	24	1.4	25	US-08-775-609-3	Sequence 3, Appli
19	24	1.4	25	US-08-775-607-3	Sequence 3, Appli
20	24	1.4	25	US-08-775-607-3	Sequence 3, Appli
21	24	1.4	30	US-08-068-747-2	Sequence 2, Appli
22	24	1.4	33	US-08-068-747-7	Sequence 7, Appli
23	24	1.4	69	US-08-480-173A-31	Sequence 31, Appli
24	24	1.4	69	US-08-480-173A-32	Sequence 32, Appli
25	24	1.4	69	US-08-484-408A-31	Sequence 31, Appli
26	24	1.4	69	US-08-484-408A-32	Sequence 32, Appli
27	24	1.4	103	US-08-480-173A-12	Sequence 12, Appli
28	24	1.4	103	US-08-484-408A-12	Sequence 12, Appli

29	24	1.4	457	2	US-08-332-766A-7	Sequence 7, Appli
30	24	1.4	519	2	US-08-581-528A-5	Sequence 5, Appli
31	24	1.4	519	4	PCT-US94-07799-5	Sequence 1, Appli
32	24	1.4	633	1	US-08-234-783-1	Sequence 1, Appli
33	24	1.4	633	1	US-08-456-907-1	Sequence 1, Appli
34	24	1.4	633	4	PCT-US95-05523-1	Sequence 29, Appli
35	24	1.4	1203	1	US-08-362-670B-29	Sequence 29, Appli
36	24	1.4	1203	3	US-08-333-576C-29	Sequence 29, Appli
37	24	1.4	1203	4	PCT-US94-14030A-29	Sequence 29, Appli
38	24	1.4	1393	2	US-08-765-875-1	Sequence 1, Appli
39	24	1.4	1393	3	US-08-795-671-1	Sequence 1, Appli
40	24	1.4	1769	2	US-08-765-268A-1	Sequence 1, Appli
41	24	1.4	1809	4	US-08-455-001-1	Sequence 1, Appli
42	24	1.4	1809	4	PCT-US95-11869-1	Sequence 1, Appli
43	24	1.4	2953	2	US-08-859-201-1	Sequence 1, Appli
44	24	1.4	3090	4	PCT-US93-06251-7	Sequence 7, Appli
45	24	1.4	3765	3	US-07-705-490-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-083-351-1  
Sequence 1, Application US/09083351  
Patent No. 6087107  
GENERAL INFORMATION:  
APPLICANT: Sheffield, Val C.  
APPLICANT: Alward, Wallace L.M.  
APPLICANT: Stone, Edwin M.  
APPLICANT: Nishimura, Darryl  
APPLICANT: Patil, Shiva  
TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,351  
FILING DATE: 22-MAY-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Arnold, Beth E.  
REGISTRATION NUMBER: 35,430  
REFERENCE/DOCKET NUMBER: UIA-029.02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-832-1000  
TELEFAX: 617-832-7000  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3946 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 475..2133  
US-09-083-351-1  
Query Match 73.2%; Score 1217; DB 3; Length 3946;

Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1657; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

```
OY 1 atgacgagcgctactcgcgtgtccagcccaactccctgagagtggtccctactcgcgc 60
Db 475 ATGAGGCGCGCTACTCCCTGTCAGCCCACTCCCTGAGAGTGTGCTTACTCCGCG 534
OY 61 ggcagacgaagctactaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120
Db 535 GCGAGGAGAGCTACTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 594
OY 121 ccggcccccaatgagcgtgtactcgcacccctgcgcgcgcgcgcgcgcgcgcgcgc 180
Db 595 CCGGCGCCCATGAGGTGCTACTCGCACCCCTGGCGCGCGCGCGCGCGCGCGCGCGCG 654
OY 181 gcccgcgcctacgagcgccctacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240
Db 655 GCGCGCGCTTACGGGCGCTTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 714
OY 241 tatagctacatcgctcatalcaccatgacacgcgcgcgcgcgcgcgcgcgcgcgcgc 300
Db 715 TATAGCTACATGCGCGCTCATCAGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 774
OY 301 ctgacgagcatctacacatctacatgagccgcctccctctacccgcgcgcgcgcgcgc 360
Db 775 CTGAGCGGCTATCCACAGTTATCATGACCGCTTCCCTTCTACCGGCGAGACACAGCG 834
OY 361 gcttgagacacacatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 420
Db 835 GCGTGCGAGACAGCATCCGCGCACACCTTCGCTCAAGAGTCTTCTGTCAGAGTCCG 894
OY 421 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 480
Db 895 CGCGACGACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 954
OY 481 atgttcagagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 955 ATGTTGAGAGAGCGCGCTTCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1014
OY 541 aaggaacagagagagagagagagagagagagagagagagagagagagagagagagag 599
Db 1015 AAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1074
OY 600 cccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 659
Db 1075 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1133
OY 660 cgttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 719
Db 1134 CGTGGCGATCCAGAGCATCAAGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1193
OY 720 gtcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 779
Db 1194 GTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1253
OY 780 cgaacagacagacagacagacagacagacagacagacagacagacagacagacagac 839
Db 1254 CGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1313
OY 840 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 899
Db 1314 GCGGCGCGCTACGCTGAGCGGTGCGGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1373
OY 900 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 959
Db 1374 GCGCGCGCGCATAGCGAGGCTTCAAGCTTGAGCAACATCATGAGTGTGCGCGCGGTC 1433
OY 960 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1019
Db 1434 GCGCGAGAGCGCGCGCGCGAGAGTCAAGCTTGTGCGCTTGTGCGCGCGCGCGCGCTC 1493
OY 1020 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1079
```

```
Db 1494 GCGCGCGGAGATCGCACCCCGCTGCGCTGCGCGCTTACTCGCGCGCGAGAGCTCCT 1553
OY 1080 ctacagctccccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1139
Db 1554 CTACAGCTCCCCCTGCGAGCGACAGATCTTCAAGCGCGGAGCTCGGGCGCGCGCGCG 1613
OY 1140 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1199
Db 1614 GCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1673
OY 1200 gacgcgttagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1259
Db 1674 GAGCTGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1733
OY 1260 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1319
Db 1734 CGGCTGCGCGGTGAGACACCCCTGCGCGAGCTTCTTGTCTTCTGCTCAGAGAGAG 1793
OY 1320 ctgcgttcctcctgagtcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1379
Db 1794 CTGCTGTCTTCTGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1853
OY 1380 ccaacacccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1439
Db 1854 CCACACACCTTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1499
OY 1440 agacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1559
Db 1914 AGACCTGCGCGCTTGTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1619
OY 1500 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1673
Db 1974 GCGCGAGAGCTTCTGCTGCGCGAGAGTGTTCAGATCAGAGAGATCGCTTGAACAA 2033
OY 1560 ctctcagtgaaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1699
Db 2034 CTCTCAGTGAACGCGGAGATGAGCTGTCAATGAGCTTCTTCTTCTTCTTCTTCTT 2093
OY 1620 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1759
Db 2094 CCGCACGTCCGAGGCTTCTGCTACGACTGAGCAAGTTTGA 2136
```

RESULT 2  
US-09-083-351-3  
; Sequence 3, Application US/09083351  
; Patent No. 6087107  
; GENERAL INFORMATION:  
; APPLICANT: Sheffield, Val C.  
; APPLICANT: Alward, Wallace L.M.  
; APPLICANT: Stone, Edwin M.  
; APPLICANT: Nishimura, Darryl  
; APPLICANT: Patil, Shiva  
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR  
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & FLIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/083,351  
; FILING DATE: 22-MAY-1998  
; CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
NAME: Arnold Beth E

```

1 REGISTRATION NUMBER: 35,430
2 REFERENCE/DOCKET NUMBER: UIA-029,022
3
4 TELECOMMUNICATION INFORMATION:
5
6 TELEPHONE: 617-832-1000
7
8 TELEFAX: 617-832-7000
9
10 INFORMATION FOR SEQ. ID NO: 3:
11
12 SEQUENCE CHARACTERISTICS:
13
14 LENGTH: 1659 base pairs
15
16 TYPE: nucleic acid
17
18 STRANDEDNESS: single
19
20 TOPOLOGY: linear
21
22 MOLECULE TYPE: cDNA
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
10
```

Query Match	73.0%;	Score 1214;	DB 3;	Length 1659;
Best Local Similarity	99.6%;	Pred. No. 0;		
Matches 1654;	Conservative	0;	Mismatches 4;	Indels 2;
			Gaps	2;

[illegible][illegible]

us-09-292-862-1.rni

RESULT 4  
PCT-US93-04102-4  
Sequence 4, Application PC/TUS9304102  
GENERAL INFORMATION:  
APPLICANT: Tao, Wufan  
APPLICANT: Lai, Eseng  
TITLE OF INVENTION: BRAIN TRANSCRIPTION FACTOR, NOCUTIC  
TITLE OF INVENTION: ACIDS ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: c/o Cooper and Dunham, 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04102  
FILING DATE: 19930430  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,292  
FILING DATE: 13-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.

```

1      RESULT      5
2      US-08-331-644-3
3      ; Sequence 3, Application US/08331644
4      ; Patent No. 5976872
5      ;
6      GENERAL INFORMATION:
7      APPLICANT:   Tao, Mufan
8      APPLICANT:   Lai, Eseng
9      TITLE OF INVENTION:  BRAIN TRANSCRIPTION FACTOR, NUCLEIC
10     TITLE OF INVENTION:  ACIDS ENCODING SAME AND USES THEREOF
11     NUMBER OF SEQUENCES:  4
12     CORRESPONDENCE ADDRESS:
13     ADDRESSEE:  Cooper & Dunham
14     STREET:    185 Avenue of the Americas
15     CITY:     New York
16     STATE:    New York
17     COUNTRY:   USA
18     ZIP:      10036
19
20     COMPUTER READABLE FORM:
21     MEDIUM TYPE:  Floppy disk
22     COMPUTER:  IBM PC compatible
23     OPERATING SYSTEM:  PC-DOS/MS-DOS
24     SOFTWARE:  Patentlin Release #1.0, Version #1.25
25     CURRENT APPLICATION DATA:
26     APPLICATION NUMBER:  US/08/331,644
27     FILING DATE:
28
29     CLASSIFICATION:  435
30     PRIOR APPLICATION DATA:
31     APPLICATION NUMBER:  US 07/882,292
32     FILING DATE:  13-MAY-1992
33     ATTORNEY/AGENT INFORMATION:
34     NAME:  White, John P.
35     REGISTRATION NUMBER:  28,678
36     REFERENCE/DOCKET NUMBER:  41472-A-PCT-US
37     TELECOMMUNICATION INFORMATION:
38     TELEPHONE:  212-278-0400
39     TELEFAX:  212-391-0525
40
41     INFORMATION FOR SEQ ID NO: 3:
42     SEQUENCE CHARACTERISTICS:
43     LENGTH:  1860 base pairs
44     TYPE:  nucleic acid
45     STRANDEDNESS:  single
46     TOPOLOGY:  linear
47
48     MOLECULE TYPE:  CDNA
49
50     US-08-331-644-3

```

Query Match	1.7%;	Score 28;	DB 2;	Length 1860;
Best Local Similarity	100.0%;	Pred. No.	0.0048;	





Query Match 1.6%; Score 27; DB 3; Length 804;  
Best Local Similarity 100.0%; Pred. No. 0.014;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1342 ggcggcgccgagcgccgagcg 1368  
|||||  
DB 28 ggcggcgccgagcgccgagcg 54

RESULT 11  
US-08-149-103-1

Sequence 1, Application US/08149103  
Patent No. 5750367

GENERAL INFORMATION:

APPLICANT: Lawrence C. B. Chan

TITLE OF INVENTION: HUMAN AND MOUSE VERY LOW

TITLE OF INVENTION: DENSITY LIPOPROTEIN RECEPTORS

TITLE OF INVENTION: AND METHODS FOR USE OF SUCH

RECEPTORS

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: LYON & LYON

STREET: 611 West Sixth Street

CITY: Los Angeles

STATE: California

COUNTRY: U.S.A.

ZIP: 90017

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: IBM MS-DOS (Version 5.0)

SOFTWARE: WordPerfect (Version 5.1)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/149,103

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

PRIOR APPLICATION DATA: including application

PRIOR APPLICATION DATA: described below:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Warburg, Richard J.

REGISTRATION NUMBER: 32,327

REFERENCE/DOCKET NUMBER: 264/052

TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440

TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3330 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-149-103-1

Query Match

Best Local Similarity 1.5%; Score 25; DB 1; Length 3330;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1337 acggcgccgagcgccgagcg 1361

|||||

DB 67 acggcgccgagcgccgagcg 91

RESULT 12

US-08-451-883-1

Sequence 1, Application US/08451883

Patent No. 5798209

GENERAL INFORMATION:

APPLICANT: Lawrence C.B. Chan

TITLE OF INVENTION: HUMAN AND MOUSE VERY LOW DENSITY

TITLE OF INVENTION: LIPOPROTEIN RECEPTORS AND METHODS FOR

TITLE OF INVENTION: USE OF SUCH RECEPTORS

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: LYON & LYON

STREET: 633 West Fifth Street, Suite 4700

CITY: Los Angeles

STATE: California

COUNTRY: U.S.A.

ZIP: 90071-2066

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: IBM MS-DOS (Version 6.22)

SOFTWARE: WordPerfect (Version 5.1)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/451,883

FILING DATE: May 26, 1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

PRIOR APPLICATION DATA: including application

PRIOR APPLICATION DATA: described below:

APPLICATION NUMBER: 08/149,103

FILING DATE: No. 5798209ember 8, 1993

ATTORNEY/AGENT INFORMATION:

NAME: Knight, Matthew W.

REGISTRATION NUMBER: 36,846

REFERENCE/DOCKET NUMBER: 212/268

TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440

TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3330 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-451-883-1

Query Match

Best Local Similarity 1.5%; Score 25; DB 1; Length 3330;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1337 acggcgccgagcgccgagcg 1361

|||||

DB 67 acggcgccgagcgccgagcg 91

RESULT 13

US-08-393-734-1

Sequence 1, Application US/08393734

Patent No. 5652224

GENERAL INFORMATION:

APPLICANT: Wilson, James M.

APPLICANT: Kozarsky, Karen F.

APPLICANT: Straus, Jerome F.

TITLE OF INVENTION: Methods and Compositions for Gene

TITLE OF INVENTION: Therapy for the Treatment of Defects in Lipoprotein

TITLE OF INVENTION: Metabolism

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Cntr., PO Box 457

CITY: Spring House

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

[illegible]

Query Match	1.5%;	Score 25;	DB 1;	Length 8438;
Best Local Similarity	100.0%;	Pred. No. 0.065;		
Matches 25; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

QY	1340	g c g g c g g c g g c g g c g g g g g	1364
Db	5651	g c g g c g g c g g c g g c g g g g g	5675

```

RESULT 15
US-08-393-734-3
; Sequence 3, Application US/08393734
; Patent No. 5652224
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Kozarsky, Karen F.
; APPLICANT: Strauss, Jerome F.
; TITLE OF INVENTION: Methods and Compositions for Gene
; TITLE OF INVENTION: Therapy for the Treatment of Defects in Lipoprotein
; TITLE OF INVENTION: Metabolism
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr., PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,734
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: UPMH1254USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9592 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown

```

















[illegible]

RESULT	5
231671	
ID	231671 standard; DNA; 3946 BP.
XX	
XX	231671;
AC	
XX	
DT	17-JAN-2000 (first entry)
XX	
DE	Human FHL7 coding sequence.
XX	
KW	FHL7; human; forkhead transcript
KW	congenital heart disease; ss.
XX	
OS	Homo sapiens.

XX WO9952415-A2.  
 PN  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 14-APR-1999; 99WO-US08159.  
 XX  
 PR 15-APR-1998; 98US-0081870.  
 PR 22-MAY-1998; 98US-0083351.  
 XX  
 PA (IOWA ) UNIV IOWA RES FOUND.  
 XX  
 XX Sheffield VC, Alward WLM, Stone EM, Nishimura D, Patil S;  
 PI  
 XX WFI; 1999-620257/53.  
 DR  
 XX P-PSDB; Y43260.  
 DR  
 XX  
 PT New isolated human forkhead transcription factor gene, FKHL7, used to  
 XX develop products for the diagnosis, prognosis, monitoring, prevention  
 PT or treatment of congenital heart disease -  
 XX  
 XX  
 PS Claim 11; Fig 1; 98pp; English.  
 XX  
 XX This sequence encodes the human forkhead transcription factor gene,  
 CC designated FKHL7, of the invention. FKHL7 can be used in a novel method  
 CC for treating or preventing the development of a congenital heart disease  
 CC (CHD) in a subject. The FKHL7 sequences can be used for diagnosis,  
 CC prognosis, monitoring, prevention and treatment of CHD. They can also  
 CC used for the production of transgenic animals and drug screening.  
 XX  
 XX Sequence 3946 BP; 905 A; 1155 C; 1053 G; 833 T; 0 other;

[illegible]

```

OY 541 aaggaagaaggaagaaggaagctgacactcaagaagagccgccccgcgg -cgacag 599
Db 1015 aaggaagaaggaagaaggaagctgacactcaagaagagccgccccgcggcag 1074
OY 600 ccccccgcggcgccgcggcgagcagcgccgaacgagccgcgctcgcagccgcgc 659
Db 1075 ccccccgcggcgccgcggcgagcagcgccgaacgagccgcgctcgcagccgcgc 1133
OY 660 cgtgcgcatccaagaacatcaagaaggaaggtacgtgccccctcgccgccccagccccct 719
Db 1134 cgtgcgcatccaagaacatcaagaaggaaggtacgtgccccctcgccgccccagccccct 1193
OY 720 gtccccgcggcgccgcgctggtgacgagcgagcgccgcggtgccccagaatcgagaagccc 779
Db 1194 gtccccgcggcgccgcgctggtgacgagcgagcgccgcggtgccccagaatcgagaagccc 1253
OY 780 cgaacagcagcagcagcagcagcctgtccagcgagcagccccccgcggcagcagcctgcctgcgc 839
Db 1254 cgaacagcagcagcagcagcagcctgtccagcgagcagccccccgcggcagcagcctgcctgcgc 1313
OY 840 gcggcgccgctcagcctgagcagctgacgattccgagcgccgcggcgccgcgccccctcgcgc 899
Db 1314 gcggcgccgctcagcctgagcagctgacgattccgagcgccgcggcgccgcgccccctcgcgc 1373
OY 900 gcgcgcgcacacatagcagagcgttcagcgctggaacacatacgaagtcgctgcgggggtc 959
Db 1374 gcgcgcgcacacatagcagagcgttcagcgctggaacacatacgaagtcgctgcgggggtc 1433
OY 960 gccgcagagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1019
Db 1434 gccgcagagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1493
OY 1020 gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1079
Db 1494 gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1553
OY 1080 ctacaagctcccccctgcagcagcagcctccagcgcgagcagcctcgcgcgcgcgcgcgcgcgcgc 1139
Db 1554 ctacaagctcccccctgcagcagcagcctccagcgcgagcagcctcgcgcgcgcgcgcgcgcgcgc 1613
OY 1140 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1219
Db 1614 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1673
OY 1200 gacgcgtgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1259
Db 1674 gacgcgtgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1733
OY 1260 cgcgcgtgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1319
Db 1734 cgcgcgtgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1793
OY 1320 ctgctgctccctgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1379
Db 1794 ctgctgctccctgaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1853
OY 1380 ccaccacccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1439
Db 1854 ccaccacccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1913
OY 1440 agaccctggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1499
Db 1914 agaccctggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1973
OY 1500 gcaagcagaactccaactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1559
Db 1974 gcaagcagaactccaactcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 2033
OY 1560 ctctccagctgaacggaatagtagctgtcaaatgacctccctccagccagcgtctctgtta 1619
Db 2034 ctctccagctgaacggaatagtagctgtcaaatgacctccctccagccagcgtctctgtta 2093
OY 1620 ccgcaagctcggagagcttctgtctaggaactgttagaagtttttga 1662

```

```

Db 2094 ccgcaagctcggagagcttctgtctaggaactgttagaagtttttga 2136
RESULT 6
X28103
ID X28103 standard; DNA: 2106 BP.
AC X28103:
DE 11-JUN-1999 (first entry)
DE Freac11 gene.
KW Freac11: transcription factor; adipose tissue expressed gene; obesity;
KW lipid metabolism; adipocyte differentiation; obesity-related condition;
KW non-insulin dependent diabetes mellitus; cardiovascular disease;
KW catabolic condition; anorexia; bulimia; therapy; ss.
OS Homo sapiens.
PN W09854216-A1.
PD 03-DEC-1998.
PF 26-MAY-1996; 98WO-SE00989.
PR 26-MAY-1997; 97SE-0001963.
PA (PHMA) PHARMACIA & UPJOHN AB.
PI Carlsson P, Enerbaeck S;
DR WPI: 1999-070211/06.
DR P-PSDB: Y01097.
PT New transcription factor designated FREAC11 - which regulates
PT adipose tissue expressed genes involved in lipid metabolism and
PT adipocyte differentiation, used to, e.g. inhibit gene expression
PS Claim 1; Fig 1; 23pp; English.
XX
XX This sequence represents the Freac11 gene of the invention. The encoded
XX protein has transcriptional regulatory function directed against adipose
XX tissue expressed genes. The protein is involved in lipid metabolism
XX and/or adipocyte differentiation. FREAC11 is selectively expressed in
XX adipose tissue, and is used for transcriptional regulation of adipocyte
XX expressed genes. The DNA or protein can be used as drugs for treating
XX obesity-related conditions to increase or decrease the activity of
XX adipocyte expressed genes, e.g. FREAC11 can be used as a drug that blocks
XX the cis elements of adipose expressed genes to inhibit gene expression;
XX antisense constructs of Freac11 can be used to down regulate expression.
XX The FREAC11 protein can also be used for high throughput screening for
XX substances that affect the activity of FREAC11, such as inhibitors,
XX antagonists, or agonists. Conditions which can be treated include
XX obesity, non-insulin dependent diabetes mellitus, cardiovascular
XX diseases, catabolic conditions, anorexia, bulimia.
XX
SQ Sequence 2106 BP; 419 A; 757 C; 602 G; 328 T; 0 other;
Query Match 4.1%; Score 68; DB 20; Length 2106;
Best Local Similarity 100.0%; Pred. No. 7e-20;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 352 aacaagcagggctgagcagaacgacatccgacacacctcgtctcaagaggtcttcgtc 411
Db 348 aacaagcagggctgagcagaacgacatccgacacacctcgtctcaagaggtcttcgtc 407
OY 412 aaggtgcc 419
Db 408 aaggtgcc 415

```

```

RESULT      7
050633
ID   Q50633 standard; cDNA; 1155 BP.
XX
AC   Q50633;
XX
DT   02-JUN-1994 (first entry)
XX
DE   Brain factor-3 partial sequence.
XX
KW   Brain factor; BF-1; BF-2; BF-3; DNA binding domain;
KM   brain transcription factor; diagnosis; tumour; cancer; probe;
XX   telencephalon; ss.
OS   Rattus rattus.
XX
PN   W09323430-A.
XX
PD   25-NOV-1993.
XX
PF   30-APR-1993;    93WO-US04102.
XX
PR   13-MAY-1992;    92US-0882292.
XX
PA   (SLOK ) SLOAN KETTERING INST CANCER.
XX
PL   Lai E, Tao W;
XX
DR   WPI; 1993-386481/48.
XX
PT   Isolated, animal nucleic acids encoding brain transcription
PT   factors - useful for diagnosis and treatment of abnormal brain
PT   factor synthesis in tumour tissue from animals and utilised as
PT   probe
XX
PS   Claim 3; Page 63; 96pp; English.
XX
CC   The brain factors are transcription factors. Abnormal expression of
CC   BF-1 in telencephalon-derived tissue or tumour tissue can be
CC   diagnosed. BF DNA or protein can be used to correct defective
CC   synthesis of BF.
XX
SQ   Sequence 1155 BP; 211 A; 414 C; 367 G; 163 T; 0 other;

Query Match          2.8%; Score 47; DB 14; Length 1155;
Best Local Similarity 100.0%; Pred. No. 5,6e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY   415 gtgcgcgcgacagacaagaagccggcgaaggcgcaactactagacct 461
      |||||||
DB   682 gtgcgcgcgacagacaagaagccggcgaaggcgcaactactagacct 728

RESULT      8
050632
ID   Q50632 standard; cDNA; 1860 BP.
XX
AC   Q50632;
XX
DT   02-JUN-1994 (first entry)
XX
DE   Brain factor-2.
XX
KW   Brain factor; BF-1; BF-2; BF-3; DNA binding domain;
KM   brain transcription factor; diagnosis; tumour; cancer; probe;
XX   telencephalon; ss.
OS   Rattus rattus.
XX
PN   W09323430-A.
XX
```

PD	25-NOV-1993.
XX	
PF	30-APR-1993; 93WO-US04102.
XX	
PR	13-MAY-1992; 92US-0882292.
XX	
PA	(SLOAN ) SLOAN KETTERING INST CANCER.
XX	
PI	Lai E, Tao W;
XX	
DR	WPI; 1993-386481/48.
XX	
PT	Isolated, animal nucleic acids encoding brain transcription
PT	factors - useful for diagnosis and treatment of abnormal brain
PT	factor synthesis in tumour tissue from animals and utilised as
PT	probe
XX	
PS	Claim 2; Page 62-63; 96pp; English.
XX	
CC	The brain factors are transcription factors. Abnormal expression of
CC	BF-1 in telencephalon-derived tissue or tumour tissue can be
CC	diagnosed. BF DNA or protein can be used to correct defective
CC	synthesis of BF.
XX	
SO	Sequence 1860 BP; 402 A; 543 C; 545 G; 370 T; 0 other;
Query Match	1.7%; Score 28; DB 14; Length 1860;
Best Local Similarity	100.0%; Pred. No. 0.0057;
Matches	28; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
QY	247 tacatcgcgctcaccacatgcatcc 274
Db	768 tacatcgcgctcaccacatgcatcc 795
RESULT	9
ID	V04426
XX	V04426 standard; cDNA; 640 BP.
XX	
AC	V04426;
XX	
DT	27-APR-1998 (first entry)
XX	
DE	Human calpain small subunit cDNA fragment.
XX	
KW	Calpain; human; leukocyte; calcium dependent cysteine protease;
KW	screening; activator; inhibitor; treatment; prevention; cancer;
KW	cerebral apoplexy; cerebral infarction; subarachnoid haemorrhage;
KW	Alzheimer's disease; myodystrophy; cataracts; collagen disease;
KW	ischemic heart disease; atherosclerosis; arthritis;
KW	small subunit; ds.
XX	
OS	Homo sapiens.
XX	
PN	EP799892-A2.
XX	
PD	08-OCT-1997.
XX	
PE	03-APR-1997; 97EP-0105508.
XX	
PR	05-APR-1996; 96JP-0083649.
XX	
PA	(TAKEDA ) TAKEDA CHEM IND LTD.
XX	
PI	Kawamoto T, Nishi K, Shintani Y;
XX	
DR	WPI; 1997-482674/45.
XX	
PT	Human calpain protein and related DNA - useful for drug screening
PT	and treating cancer, stroke, etc.
XX	
XX	Example 2; Page 38; 43pp; English.











Wed Feb 7 14:05:53 2001

us-09-292-862-1.rng

---

Page 12

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 6, 2001, 18:01:15 : Search time 1965.59 Seconds  
(without alignments)  
4327.302 Million cell updates/sec

Title: US-09-292-862-1  
Perfect score: 1662  
Sequence: 1 atgcagcgcgcgtactccgt.....acgactctagcaagtttga 1662

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1118133 seqs, 2558875100 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl:  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_ph:\*  
6: gb\_pl1:\*  
7: gb\_pl2:\*  
8: gb\_pr1:\*  
9: gb\_pr2:\*  
10: gb\_pr3:\*  
11: em\_fun:\*  
12: em\_hum1:\*  
13: em\_hum2:\*  
14: em\_in:\*  
15: em\_om:\*  
16: em\_or:\*  
17: em\_ov:\*  
18: em\_pat:\*  
19: em\_ph:\*  
20: em\_pl:\*  
21: em\_ro:\*  
22: em\_sts:\*  
23: em\_sy:\*  
24: em\_un:\*  
25: em\_vl:\*  
26: gb\_hlg1:\*  
27: gb\_hlg2:\*  
28: gb\_in1:\*  
29: gb\_in2:\*  
30: em\_ba1:\*  
31: em\_ba2:\*  
32: em\_hum3:\*  
33: em\_hum4:\*  
34: gb\_pr4:\*  
35: gb\_hlg3:\*  
36: gb\_hlg4:\*  
37: gb\_hlg5:\*  
38: gb\_hlg6:\*  
39: gb\_hlg7:\*  
40: em\_hlg1:\*  
41: em\_hlg2:\*  
42: em\_hlg3:\*  
43: em\_hum5:\*

44: gb\_pl3:\*  
45: gb\_pr5:\*  
46: gb\_hlg8:\*  
47: gb\_hlg9:\*  
48: gb\_hlg10:\*  
49: gb\_hlg11:\*  
50: gb\_hlg12:\*  
51: gb\_hlg13:\*  
52: gb\_hlg14:\*  
53: gb\_in3:\*  
54: gb\_hlg15:\*  
55: gb\_hlg16:\*  
56: gb\_hlg17:\*  
57: em\_hlg4:\*  
58: em\_hlg5:\*  
59: em\_hlg6:\*  
60: em\_hlg7:\*  
61: em\_hum6:\*  
62: gb\_hlg18:\*  
63: gb\_hlg19:\*  
64: gb\_ba3:\*  
65: em\_hlg8:\*  
66: em\_hlg9:\*  
67: em\_hlg10:\*  
68: gb\_pr6:\*  
69: gb\_pr7:\*  
70: gb\_hlg20:\*  
71: gb\_hlg21:\*  
72: gb\_hlg22:\*  
73: gb\_hlg23:\*  
74: gb\_ro:\*  
75: gb\_sts1:\*  
76: gb\_sts2:\*  
77: gb\_sy:\*  
78: gb\_un:\*  
79: gb\_vl1:\*  
80: gb\_vl2:\*  
81: gb\_pat1:\*  
82: gb\_pat2:\*  
83: em\_hlg0:\*  
84: gb\_hlg24:\*  
85: gb\_pr8:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1662	100.0	1662	10	AF078096	AF078096 Homo sapi
2	1319	79.4	104729	45	HS118B18	AL034344 Human DNA
3	1217	73.2	3946	10	AF048693	AF048693 Homo sapi
4	986	59.3	160855	35	AC011579	AC011579 Homo sapi
5	318	19.1	318	69	HSU13221	U13221 Human forch
6	236	14.2	331	85	HOMFORKHOC	L12143 Human Droso
7	152	9.1	249	44	HS116B11F	Z56386 H. sapiens C
8	74	4.5	334	74	MMFKH1A	X71939 M. musculus
9	74	4.5	355	74	MUSPRKHA	L10406 Mouse fork
10	74	4.5	3579	74	AF045017	AF045017 Mus muscu
11	74	4.5	3637	74	HSMPFH1	Y08223 H. sapiens M
12	68	4.1	3289	69	AC009108	AC009108 Homo sapi
13	68	4.1	170431	27	RATHFH3BF	L13193 Rattus norv
14	48	2.9	303	74	AR083461	AR083461 Sequence
15	47	2.8	1155	81	SG3607	SG3607 MFR-1-mesen
16	47	2.8	1942	74	MMKRFH1	X92499 M. musculus
17	47	2.8	2323	74	MMKRFH1	X74040 M. musculus
18	47	2.8	2712	74	MMKRFH1	Y08222 M. musculus
19	47	2.8	6021	74	MMKRFH1	U13225 Human forch
20	45	2.7	318	69	HSU13225	U13225 Human forch
21	42	2.5	180	4	GU037277	U37277 Gallus gall





repeat_region	/note="Charlie2 repeat: matches 3390. .3621 of consensus 5089. .5465
repeat_region	/note="MER69 repeat: matches 1. .410 of consensus" 5678. .5743
repeat_region	/note="Charlie2 repeat: matches 3636. .3668 of consensus" 5744. .5971
repeat_region	/note="AluSc repeat: matches 65. .292 of consensus" 5972. .6053
repeat_region	/note="Charlie2 repeat: matches 3668. .3761 of consensus" 6166. .6491
repeat_region	/note="HERV_L repeat: matches 2050. .2362 of consensus" 6492. .6801
repeat_region	/note="AluYb8 repeat: matches 1. .318 of consensus" 6802. .7299
repeat_region	/note="HERV_L repeat: matches 2362. .2905 of consensus" 7412. .7455
repeat_region	/note="11 copies 4 mer tdtg 95% conserved" 7520. .7828
repeat_region	/note="AluJo repeat: matches 1. .300 of consensus" 8051. .8276
misc_feature	/note="HERV_L repeat: matches 5135. .5369 of consensus" 9107. .9568
misc_feature	/note="match: GSS: Em:AQ261001" 9267. .10109
repeat_region	/note="match: GSS: Em:AQ742778" 10378. .10681
misc_feature	/note="AluSg repeat: matches 1. .300 of consensus" 11116. .11502
repeat_region	/note="match: GSS: Em:AQ135654" 11803. .11979
repeat_region	/note="FRAM repeat: matches 1. .176 of consensus" 12020. .12175
repeat_region	/note="MR repeat: matches 88. .262 of consensus" 12453. .12757
repeat_region	/note="AluSx repeat: matches 1. .308 of consensus" 13029. .13056
repeat_region	/note="7 copies 4 mer caaa 92% conserved" 13193. .13564
repeat_region	/note="MSTA repeat: matches 6. .426 of consensus" 13612. .13896
repeat_region	/note="AluSx repeat: matches 1. .285 of consensus" 13900. .14466
repeat_region	/note="LIMC5 repeat: matches 7315. .7903 of consensus" 14625. .14796
repeat_region	/note="MR repeat: matches 20. .208 of consensus" 14820. .14946
repeat_region	/note="MSTD repeat: matches 217. .342 of consensus" 15065. .15209
repeat_region	/note="MSTD repeat: matches 1. .160 of consensus" 15839. .16043
repeat_region	/note="MR repeat: matches 24. .234 of consensus" 16269. .16474
repeat_region	/note="L2 repeat: matches 2383. .2601 of consensus" 16800. .16991
repeat_region	/note="LIMD2 repeat: matches 5732. .5927 of consensus" 17104. .17171
repeat_region	/note="L2 repeat: matches 2652. .2750 of consensus" 18948. .19266
repeat_region	/note="AluY repeat: matches 1. .311 of consensus" 19348. .20141
repeat_region	/note="LIMC repeat: matches 1064. .1854 of consensus" 20142. .20449
repeat_region	/note="AluJo repeat: matches 1. .312 of consensus" 20450. .20625
repeat_region	/note="LIMC repeat: matches 1854. .2029 of consensus" 20687. .21037
repeat_region	/note="LIM4 repeat: matches 4389. .4752 of consensus" 21063. .21181
repeat_region	/note="AluSg/x repeat: matches 2. .133 of consensus" 21207. .21508
repeat_region	/note="AluSp repeat: matches 1. .303 of consensus" 21556. .21728
repeat_region	/note="LIM1 repeat: matches 2216. .2385 of consensus"

```

repeat_region 21729. .22032
             /note="Aluob repeat: matches 3. .306 of consensus"
repeat_region 22033. .22776
             /note="L1M1 repeat: matches 1076. .2216 of consensus"
repeat_region 22782. .23114
             /note="L1M1 repeat: matches 2513. .3695 of consensus"
repeat_region 24315. .24403
             /note="L1P45 repeat: matches 6053. .6141 of consensus"
repeat_region 24404. .24961
             /note="L1M1 repeat: matches 1971. .2513 of consensus"
repeat_region 25142. .27426
             /note="L1MC repeat: matches -268. .1596 of consensus"
repeat_region 27583. .27951
             /note="L1T16 repeat: matches 143. .511 of consensus"
misc_feature 27921. .28165
             /note="match: STS: Em:G02957; match: STS: Em:G02957"
misc_feature 27998. .28164
             /note="match: STS: Em:G59738"
repeat_region 29617. .29696
             /note="MIR repeat: matches 60. .140 of consensus"
repeat_region 36373. .36583
             /note="L1ME repeat: matches 5580. .5799 of consensus"
repeat_region 36669. .36967
             /note="Alusg repeat: matches 1. .299 of consensus"
repeat_region 37203. .37351
             /note="L1MC/D repeat: matches 4410. .4556 of consensus"
repeat_region 37332. .37653
             /note="AlusX repeat: matches 1. .301 of consensus"
repeat_region 37654. .38270
             /note="L1MC/D repeat: matches 4556. .5186 of consensus"
repeat_region 38279. .38543
             /note="L1MC5 repeat: matches 7627. .7937 of consensus"
repeat_region 38577. .38940
             /note="L1MC/D repeat: matches 5175. .5398 of consensus"
repeat_region 41360. .41671
             /note="Aluio repeat: matches 1. .311 of consensus"
repeat_region 41994. .42081
             /note="22 copies 4 mer 1gtg 63% conserved"
repeat_region 41995. .42082
             /note="44 copies 2 mer gt 64% conserved"
repeat_region 42083. .42130
             /note="24 copies 2 mer tg 75% conserved"
repeat_region 43018. .43331
             /note="AlusX repeat: matches 1. .312 of consensus"
repeat_region 43334. .43412
             /note="12 repeat: matches 2622. .2708 of consensus"

```

[illegible]







All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: 13\_F\_18

Center clone name: 13\_F\_18

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 136346 bases at least Q40

Consensus quality: 148591 bases at least Q30

Consensus quality: 153141 bases at least Q20

Insert size: 166000; agarose-fp

Insert coverage: 155655; sum-of-ctgls

Quality coverage: 3.1 in Q20 bases; sum-of-ctgls

Quality coverage: 3.3 in Q20 bases; sum-of-ctgls

NOTE: This is a 'working draft' sequence. It currently consists of 53 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1      4649: contig of 4649 bp in length
*      4650 4749: gap of 100 bp
*      4750 5802: contig of 1053 bp in length
*      5803 5902: gap of 100 bp
*      5903 7063: contig of 1161 bp in length
*      7064 7163: gap of 100 bp
*      7164 8774: contig of 1611 bp in length
*      8775 8874: gap of 100 bp
*      8875 10112: contig of 1238 bp in length
*      10113 10212: gap of 100 bp
*      10213 11283: contig of 1071 bp in length
*      11284 11383: gap of 100 bp
*      11384 12607: contig of 1224 bp in length
*      12608 12707: gap of 100 bp
*      12708 14285: contig of 1578 bp in length
*      14286 14385: gap of 100 bp
*      14386 15414: contig of 1029 bp in length
*      15415 15514: gap of 100 bp
*      15515 17693: contig of 2179 bp in length
*      17694 17793: gap of 100 bp
*      17794 18837: contig of 1044 bp in length
*      18838 18937: gap of 100 bp
*      18938 20536: contig of 1599 bp in length
*      20537 20636: gap of 100 bp
*      20637 35792: contig of 15156 bp in length
*      35793 35892: gap of 100 bp
*      35893 37876: contig of 1984 bp in length
*      37877 37976: gap of 100 bp
*      37977 40070: contig of 2094 bp in length
*      40071 40170: gap of 100 bp
*      40171 41423: contig of 1253 bp in length
*      41424 41523: gap of 100 bp
*      41524 42814: contig of 1291 bp in length
*      42815 42914: gap of 100 bp
*      42915 44991: contig of 2077 bp in length
*      44992 45091: gap of 100 bp
*      45092 46704: contig of 1613 bp in length
*      46705 46804: gap of 100 bp
*      46805 49273: contig of 2469 bp in length
*      49274 49373: gap of 100 bp
*      49374 51528: contig of 2155 bp in length
*      51529 51628: gap of 100 bp

```

```

*      51629 52804: contig of 1176 bp in length
*      52805 52904: gap of 100 bp
*      52905 55108: contig of 2204 bp in length
*      55109 55208: gap of 100 bp
*      55209 57532: contig of 2324 bp in length
*      57533 57632: gap of 100 bp
*      57633 59640: contig of 2008 bp in length
*      59641 59740: gap of 100 bp
*      59741 61382: contig of 1642 bp in length
*      61383 61482: gap of 100 bp
*      61483 64104: contig of 2622 bp in length
*      64105 64204: gap of 100 bp
*      64205 66859: contig of 2655 bp in length
*      66860 66959: gap of 100 bp
*      66960 68324: contig of 1365 bp in length
*      68325 68424: gap of 100 bp
*      68425 71212: contig of 2788 bp in length
*      71213 71312: gap of 100 bp
*      71313 74324: contig of 3012 bp in length
*      74325 74424: gap of 100 bp
*      74425 76988: contig of 2564 bp in length
*      76989 77088: gap of 100 bp
*      77089 78929: contig of 1841 bp in length
*      78930 79029: gap of 100 bp
*      79030 81494: contig of 2465 bp in length
*      81495 81594: gap of 100 bp
*      81595 82738: contig of 1144 bp in length
*      82739 82838: gap of 100 bp
*      82839 86549: contig of 3711 bp in length
*      86550 86649: gap of 100 bp
*      86650 90087: contig of 3438 bp in length
*      90088 90187: gap of 100 bp
*      90188 93458: contig of 3271 bp in length
*      93459 93558: gap of 100 bp
*      93559 97847: contig of 4289 bp in length
*      97848 97947: gap of 100 bp
*      97948 101934: contig of 3987 bp in length
*      101935 102034: gap of 100 bp
*      102035 106306: contig of 4272 bp in length
*      106307 106406: gap of 100 bp
*      106407 110682: contig of 4276 bp in length
*      110683 110782: gap of 100 bp
*      110783 115744: contig of 4962 bp in length
*      115745 115844: gap of 100 bp
*      115845 117155: contig of 1311 bp in length
*      117156 117255: gap of 100 bp
*      117256 123338: contig of 5983 bp in length
*      123339 123338: gap of 100 bp
*      123339 128170: contig of 4832 bp in length
*      128171 128270: gap of 100 bp
*      128271 133489: contig of 5219 bp in length
*      133490 133589: gap of 100 bp
*      133590 138250: contig of 4661 bp in length
*      138251 138350: gap of 100 bp
*      138351 144214: contig of 5864 bp in length
*      144215 144314: gap of 100 bp
*      144315 151760: contig of 7446 bp in length
*      151761 151860: gap of 100 bp
*      151861 158782: contig of 6922 bp in length
*      158783 158882: gap of 100 bp
*      158883 160198: contig of 1316 bp in length
*      160199 160298: gap of 100 bp
*      160299 160855: contig of 557 bp in length.

```

#### FEATURES

##### source

1. 160855

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="RP11-13F18"

/clone\_lib="RP11 Human Male BAC"

1. 4649

/note="assembly\_fragment"

clone\_end:Sp6

vector\_side:left"

##### misc\_feature

Query Match	19.1%;	Score 318;	DB 69;	Length 318;
Best Local Similarity	100.0%;	Pred. No. 1.3e-125;		
Matches 318;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0

QY 217 cccaaggaactgttagagccgccctatgctacacatgcgcctcatcacacatgycatccag 276  
|||||

Db 1 CCCAGGACATGTTGAAGCCGCTATAGCTACATCGCGCTCATCATGCGCATCCAG 60

Qy 277 aagcccccgaagaagatcaccttaacgcatctaccagttcatctgaacgcttc 336  
|||||  
Db 61 AACGCCCGGACAGAAATCATCCCTGAGGCGCATCTACAGATTATCATGAGCCGCTTC 120  
|||||

Qy 337 cccttaccggaacaacagagctgtgacagacagcatccgcacaacctctgcctc 396  
|||||  
Db 121 CCTTCTACCGGACACAGAGAGGCTGGCAGAACAGCATCCGCCACAACCTCTGCTC 180  
|||||

Qy 397 aacgagctctgtcaaggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 456  
|||||  
Db 181 AACGATGCTTGTCTCAAGTGGCGCGACGACAGAAAGCCGGGCAAGGCGAGTACTG 240  
|||||

Qy 457 agcgtgacccgactcctacacatgtctgaagaagcagcttcctgtgcgcgcgcgcgc 516  
|||||  
Db 241 ACGGTGACCGGACTCTCAACATGTTCCAGAACGCGACCTTCTGCGCGCGCGCG 300  
|||||

Qy 517 cgcttcaagaagaagac 534  
|||||  
Db 301 CGCTTCAAGAAGAGAC 318  
|||||

RESULT 6  
HUMFORRHOC 331 bp mRNA PRI 04-AUG-1993  
LOCUS Human Drosophila forkhead homologue mRNA.  
DEFINITION L12143  
ACCESSION L12143.1 GI:347525  
VERSION  
KEYWORDS forkhead homolog.  
SOURCE Homo sapiens (library: lambda gt11) blood cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 331)  
Hromas, R.A., Moore, J.L., Johnston, T.C., Socha, C. and Klemenz, M.J.  
Drosophila forkhead homologues are expressed in a  
lineage-restricted manner in human hematopoietic cells  
JOURNAL Blood 81, 2854-2859 (1993)  
MEDLINE 93271467  
FEATURES  
location/Qualifiers  
source 1..331  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_type="chronic myelogenous leukemia"  
/tissue\_type="blood"  
/tissue\_1lb="lambda gt11"  
BASE COUNT 85 a 113 c 87 g 46 t  
ORIGIN

Query Match 14.2%; Score 236; DB 85; Length 331;  
Best Local Similarity 99.7%; Pred. No. 1.7e-90;  
Matches 286; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 232 aagccgctatagctacatcgctcatcaccatggtccatccgaagcccgagacaag 291  
|||||  
Db 31 AAGCGGCTATAGCTACATCGCGCTCATCATGCGCATCCAGAACGCCCGGACAG 90  
|||||

Qy 292 aagatccctgaagagcatctaccagttcatctgaacgcttcctcttaccgagac 351  
|||||  
Db 91 AAGATCACCCAGAACGGCATCTACAGATTCATCATGAGCGCTTCCCTTACCGGAGC 150  
|||||

Qy 352 aacaagcagagctgtgcagacacagcatcgcgcaaacctctgtctaaagaggtctgtc 411  
|||||  
Db 151 AACAGACAGGGCTGGCAGAACACATCCGCCACACACCTCTCTCAAGAGTGTCTGTC 210  
|||||

Qy 412 aaggtgc 471  
|||||  
Db 211 AAGGTGCGCGCGACGACAGAACGCCCGGCAAGGCGAGCTACTGAGACGCTGGACCCGGAC 270  
|||||

Qy 472 tccctacaacatctcgagaacgagcagcttctctgcgcgcgcgcgcgcgcgcgcgcgc 518  
|||||

Db 271 TCCACAGACATGTTGAGAACGCGAGCTTCTCGCGCGCGCGCGCGCGCGCG 317

RESULT 7  
HS116B11F/c 249 bp DNA PRI 18-OCT-1995  
LOCUS HS116B11F/c  
DEFINITION H.sapiens Cpg Island DNA genomic MseI fragment, clone 116b11,  
forward read cpg116b11.f1ta.  
ACCESSION 256386  
VERSION 256386.1 GI:1027617  
KEYWORDS Cpg island; genomic MseI fragment.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 249)  
Macdonald, M., Huckle, E., Wilkinson, P. and Micklem, G.  
Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,  
Direct Submission  
JOURNAL CB10 1RQ, England, E-mail contact: humquery@sanger.ac.uk  
TITLES 2 (bases 1 to 249)  
Cross, S.H., Charlton, J.A., Nan, X. and Bird, A.P.  
Purification of Cpg Islands using a methylated DNA binding column  
Nat. Genet. 6 (3), 236-244 (1994)  
MEDLINE 94282070  
COMMENT Vector: pGEM-5Zf(-)  
Clones are available from the UK MRC Human Genome Mapping Project  
Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:  
http://www.hgmp.mrc.ac.uk/ for details  
or contact: biophel@hgmp.mrc.ac.uk.

FEATURES  
source 1..249  
location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/sex="male"  
/dev\_stage="adult"  
/tissue\_type="blood"  
/clone\_1lb="CGI-1"  
/clone="116b11"  
BASE COUNT 48 a 52 c 66 g 81 t 2 others  
ORIGIN

Query Match 9.1%; Score 152; DB 45; Length 249;  
Best Local Similarity 100.0%; Pred. No. 1.9e-54;  
Matches 152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1511 tcaactggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1570  
|||||  
Db 249 TCCACTCGGTGCGGAGATGTTGAGTCAAGAGAGTGGCTTGAACAACTCCAGTGA 190  
|||||

Qy 1571 acggaatagtagctgtcaaatgagcttcctccatcagcagctctctgtaccgagctgcg 1630  
|||||  
Db 169 ACGGGAATAGTACTGTCAATATGAGCTTCCCTTCCAGCGAGTCTGTACCGACGCTCG 130  
|||||

Qy 1631 gagcttctgtctacagactgtgacagatttga 1662  
|||||  
Db 129 GAGCTTGTCTACGACTGTGCAAGTTTGA 98  
|||||

RESULT 8  
MERKHA 334 bp DNA ROD 25-AUG-1995  
LOCUS MERKHA  
DEFINITION M.musculus fkh-1 gene.  
ACCESSION X71939  
VERSION X71939.1 GI:311736  
KEYWORDS DNA binding domain; fork head domain.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 334)  
Kaesner, R.H.

```

/note="fork head domain; putative"
/codon_start=3
/product="fork head related protein"
/protein_id="AA03139.1"
/db_xref="GI:410056"
/translation="PPDKRWKPPYSYIALTMAIONAPDKITLNCYOFIMDRPEF
YDNKQGNQNSIRHNLSTNCECFVVPDRDKKPKQSGSYWLLDPDSYHMEFNSFLRRRR
RKKKDKAKETTFIG"

```

	Query Match	4.5%;	Score 74:	DB 74:	Length 355:
	Best Local Similarity	100.0%;	Pred. NO. 4.5e-21;		
	Matches 74;	Conservative	0;	Mismatches 0;	Indels 0; Gaps 0;
OY	463 gaccggactccacagacatgctgaagaagcgacgtcctcctgcgcygcygcgcgcgttc 522 				
Dd	255 GACCGACTCTTACACATGTTCCAGAACGGCAGCTTCGTGGCGCGCGCGCGCTTC 314 				
OY	523 aaagaagaaggaagc 536 				
Dd	315 AAGAAGAAGGACGC 328 				

RESULT	10			
LOCUS	MMAJ3298			
DEFINITION	MMAJ3298	3579 bp	DNA	ROD
ACCESSION	Mus musculus	Fkh1/Mtl	gene, complete CDS.	
VERSION	AJ223298	.1	GI:3805940	
KEYWORDS	Fkh1/Mtl	gene; transcription	factor.	
SOURCE	house mouse.			
ORGANISM	Mus musculus			

REFERENCE	Orthognathus; Muridae; Murinae; Mus.
AUTHORS	Hiemisch, H.
TITLE	Direct Substition
JOURNAL	Submitted 26-JAN-1998
REFERENCE	Bioscience AG, Im Neuenheimer Feld 519, D-69120 Heidelberg, GERMANY
AUTHORS	2 (bases 1 to 3579)
TITLE	Hiemisch, H., Schütz, G. and Keastner, K. H.
JOURNAL	The mouse Pknox1 gene: cDNA sequence, chromosomal localization
MEDLINE	and expression in adult tissues
FEATURES	Gene 220 (1-2), 77-82 (1998)
	98440371
	Location/Ortho1ifiers

source	gene	promoter	CDS
1. 3579 /organism="Mus musculus" /db_xref="taxon:10090" 1. 816 /organism="Mus musculus" /db_xref="taxon:10090" /chromosome="13" /dev_stage="embryonic day 10 p.c." /map="17.02cM" /tissue_type="embryo" 1. 2161 /gene="Fkhl/Mf1" <1. 159 /gene="Fkhl/Mf1" 500. 2161 /gene="Fkhl/Mf1" /function="putative transcription factor" /codon_start=1 /product="Fkhl/Mf1 protein"			

SVYSHPAHAEDY PMSARAATPYTOPQPDWMPY SYIALITMAIONAPDKITLNL  
GJYOFINDRFPYEDKOGMONS IKNHLSTNECEVVKPPDKRGKSGYNTDPSYNS  
MEENGSTLRRRRRFRKKDAVDKEEGRLTHQEDPPPPQAGROGAPAPPEQAGSABGP

BASE COUNT	823 a	1030 c	985 g	799 t
ORIGIN				
Query Match	4.5%;	Score 74;	DB 74;	Length 3637;
Best Local Similarity	100.0%;	Pred. No. 2,5e-21;		
Matches 74;	Conservative	0;	Mismatches 0;	Indels 0;
				Gaps 0;
Qy	463	gacccgacactctaacacatgctcagaaacgcagcttcctctgagcgagcgagcgagctc	522	
Db	1012	GACCCGACACTCTTCAACATGTTGCAAGACGACGACTTCTGCGGCGGCGGCGGCTTC	1071	
Qy	523	aagaagaagcagc	536	
Db	1072	AAGAAGAAGCAGCGC	1085	
RESULT 12				
HSMPH1		3289 bp	DNA	PRI
LOCUS	HSMPH1			14-MAY-1997
DEFINITION	H.sapiens MFH-1 gene.			
ACCESSION	Y08223			
VERSION	Y08223.1	GI:18669804		
KEYWORDS	mesenchyme fork head-1 protein; MFH-1 gene.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;			
ATTORNS	Primates; Catarrhini; Hominoideae; Homo.			
TITLE	1 (bases 1 to 3289)			
	Miura, N., Iida, K., Kakinuma, H., Yang, X.L. and Sugiyama, T.			
	Isolation of the mouse (MFH-1) and human (FKHL14) mesenchyme fork			
	head-1 genes reveals conservation of their gene and protein			
	structures			
JOURNAL	Genomics 41 (3), 489-492 (1997)			
MEDLINE	97312712			
REFERENCE	2 (bases 1 to 3289)			
AUTHORS	Miura, N.			
TITLE	Direct Submission			
JOURNAL	Submitted (18-SEP-1996) N. Miura, Akita University School of			
	Medicine, Department of Biochemistry, 1-1-1 Honcho, Akita 010, JAPAN			
	Location/Qualifiers			
FEATURES	1..3289			
source	/organism="Homo sapiens"			
gene	/db_xref="taxon:9606"			
CDS	1197..2702			
	/gene="MFH-1"			
	1197..2702			
	/gene="MFH-1"			
	/codon_start=1			
	/product="Mesenchyme Fork Head-1"			
	/protein_id="CAA69400.1"			
	/db_xref="GI:1869805"			
	/db_xref="SWISS-PROT:C99958"			
	/translation="MOARYSVDPNALGVVPILSRQNYRAAGSYGNASPMGYVSGH			
	PEQISAGKGSYAPVHHQPAKGLVKKPYEYIALITMAIONAEKKTITNGLYOTF			
	MDRRPYREKNOGQNSIRHNLISLNCFFKVPDRDKPKGSAWTLDDPSVNIKSEAA			
	FLRRRRKKKDKVSKKEEKRAHLKEPRPAASGAPATPLADAPKEAKKLVKSEAA			
	SPALPVITKVELTSPESALOGSPRSASTPAGSPDGSIPENHAAADNGLPGSVENII			
	TLRTPSGGELSPGAGRAGLVPPPLAIFYAAAPRAYGQPCAGILEAAGAGYCSM			
	AMSLYTGAEPAHICVPPALDEALSDHPSCGPTSLALNLVLAGDQSAALAAHSHNH			
	GHNDHAPPPAPQOPQPTPOPCGAAQAASWYLNHSGDLNLPGHTPAQDQITPRNY			
	REMRNSRLGIENSTIGESQVNASQCLPYRSTPLPYRHAAPVSYDCKY"			
BASE COUNT	639 a	925 g	600 t	
ORIGIN	1125 c			

```

Query Match
  4.1%; Score 68; DB 69; Length 3289;
Best Local Similarity 100.0%; Pred. No. 9.6e-19;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 aacaagcaggtctgcagacagcattccgacacacccctgcctcaagatgcttcgc 411
      |||
Db 1530 AACAAAGAGGCTGCGAGAACAGCATCCGCCAACACTCTGCTCAACGAGATGCTTCGTC 1589
      |||

QY 412 aaggtgccc 419
      |||
Db 1590 AAGGTGCC 1597

RESULT 13
LOCUS AC009108 170431 bp DNA HTG 02-SEP-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-46309, WORKING DRAFT
ACCESSION AC009108
VERSION AC009108.8 GI:9964740
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 16
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 170431)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission.
REFERENCE Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Sep 2, 2000 this sequence version replaced gi:8575957.
-----
Center: Joint Genome Institute
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 584502
Center clone name: RP11-46309
-----
Summary Statistics
Consensus quality: 167345 bases at least Q40
Consensus quality: 169423 bases at least Q30
Consensus quality: 169900 bases at least Q20
Estimated insert size: 181440; agarose-fp estimation
Estimated insert size: 170331; sum-of-contigs estimation
Quality coverage: 10.63 in Q20 bases; agarose-fp estimation
Quality coverage: 11.33 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 24820: contig of 24820 bp in length
24821 24920: gap of unknown length
37862 37862: contig of 12942 bp in length
37863 37962: gap of unknown length
114671 114671: contig of 76709 bp in length
114772 114772: gap of unknown length
114772 170431: contig of 55660 bp in length.
Location/Qualifiers
1. 170431
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"

Query Match
  4.1%; Score 68; DB 69; Length 3289;
Best Local Similarity 100.0%; Pred. No. 9.6e-19;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 352 aacaagcaggtctgcagacagcattccgacacacccctgcctcaagatgcttcgc 411
      |||
Db 1530 AACAAAGAGGCTGCGAGAACAGCATCCGCCAACACTCTGCTCAACGAGATGCTTCGTC 1589
      |||

QY 412 aaggtgccc 419
      |||
Db 1590 AAGGTGCC 1597

RESULT 14
LOCUS RATHFH3BF 303 bp mRNA ROD 26-JUL-1993
DEFINITION Rattus norvegicus brain factor-3 (HFH-BF-3) mRNA, complete cds.
ACCESSION U13193
VERSION U13193.1 GI:310156
KEYWORDS forkhead homolog; hepatocyte nuclear factor; transcription factor.
SOURCE Rattus norvegicus (strain Sprague-Dawley) adult cDNA to mRNA.
ORGANISM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
REFERENCE Rattus.
AUTHORS 1 (bases 1 to 303)
TITLE Lal,E., Prezioso,V.R., Tao,W., Chen,W.S. and Darnell,J.E.Jr.
JOURNAL Hepatocyte nuclear factor 3a belongs to a gene family in mammals
AUTHORS that is homologous to the Drosophila homeotic gene fork head
JOURNAL Unpublished (1992)
REFERENCE 2 (bases 1 to 303)
AUTHORS Clevidence,D.E., Overdier,D.G., Tao,W., Qian,X., Pani,L., Lal,E.
TITLE and Costa,R.H.
JOURNAL Identification of nine tissue-specific transcription factors of the
AUTHORS hepatocyte nuclear factor 3/forkhead DNA-binding domain family
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90, 3948-3952 (1993)
93248207
Location/Qualifiers
1. 303
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/dev_stage="adult"
<1..>303
/standard_name="HFH-BF-3"
/codon_start=1
/function="DNA binding domain"
/product="brain factor-3"
/protein_id="AAA41320.1"
/db_xref="GI:310157"
/translation="LVKPPSYVALSTMAIONAEKKITLNGIYQFIMDRFPYRENN
OQWNSIRHNLSINCEPKYVRDPRDKPKKSGSYWTLDPDSYNNFENGSLRRRRRFFK"
BASE COUNT
76 a 94 c 84 g 49 t
ORIGIN

Query Match
  2.9%; Score 48; DB 74; Length 303;
Best Local Similarity 100.0%; Pred. No. 6.6e-10;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 481 atgttcgaagcagcagcttcctgcgagcgcgcgcgcgttcaagaag 528
      |||
Db 256 ATGTTGAGAAAGCGAGCTTCTGCGCGCGCGCGCGCTTCAAGAG 303
      |||

RESULT 15

```



